

文系学生における理数系科目不安の実態把握と その解消を促進する教授法の開発

(研究課題番号 17500584)

平成17年度～平成18年度科学研究費補助金(基盤研究(C))
研究成果報告書

平成19年3月

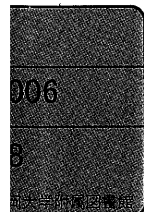
CD-ROM付き 研究代表者 山田 文康

静岡大学附属図書館

(静岡大学 情報学部)



000750932 6



【目次】

研究組織・研究成果等一覧

第1部 統計学不安尺度の構成	1
第1章 調査票の作成	1
1. 1 調査の目的	2
1. 2 先行研究の検討	3
1. 3 新たな統計学不安尺度の枠組み	4
1. 4 分野別項目の詳細	5
1. 5 アンケート票	9
第2章 調査の実施と解析	11
2. 1 調査の実施	11
2. 2 分析結果	11
2. 3 まとめ	18
第2部 統計的推測を学習するための	20
標本抽出実験シミュレーション・システム	
第1章 システムの概要	21
1. 1 目的	21
1. 2 システムの構成	22
1. 3 システムの導入	24
第2章 各学習プログラムでの課題と実施手順	26
2. 1 標本抽出実験	27
2. 2 区間推定・統計的仮説検定のためのツール	52



【研究組織】

研究代表者 山田 文康 (静岡大学 情報学部情報社会学科・教授)

研究分担者 漁田 武雄 (静岡大学 情報学部情報社会学科・教授)

高橋 晃 (静岡大学 情報学部情報社会学科・准教授)

新谷 誠 (静岡大学 情報学部情報科学科・准教授)

研究協力者 加藤 絢子 (静岡大学 情報学部情報社会学科 4 年)

【研究経費】

平成 1 7 年度 1 3 0 万円

平成 1 8 年度 1 2 0 万円

計 2 5 0 万円

【研究成果】

加藤絢子 (2006)、文系学生における統計学に対する不安意識の解析、平成 17 年度静岡大学情報学部卒業研究

山田文康・山本さつき(2007)、授業評価アンケートに対する学生の意識と評価結果

—— 学生の授業評価の妥当性と信頼性 ——、静岡大学「情報学研究」、13 卷

山田文康・山田耕史(2007)、3 教育プログラム制のもとでの入学者の学力特性

—— 入試成績と学内成績に基づいて ——、平成 18 年度入学者選抜方法研究委員会報告書

第 1 部

統計学不安尺度の作成

第1章 調査票の作成

1. 1 調査の目的

本研究は、統計学不安尺度を作成し、それを用いて、大学生が持つ統計学の学習に対する不安の実態を把握し、それが形成される要因を探ることを目的としている。

統計学不安とは、統計学を学習する場面を避けようとしたり、学習内容に対して自信を持っていないでいたりする意識を言う。例えば、社会科学や心理学といった学問を学ぶ文系学生は統計学の技能を必要とする。しかし、統計学は数式や図表の読み取りなど理数的側面を持つため、文系学生は困難に直面する。その困難やそこから生じる苦手意識の程度を測るための尺度を統計学不安尺度という。

静岡大学情報学部は、文工融合の教育を実践している。そのため、文系学生が統計学のような理数系要素を多分に含む分野を学ぶ機会が多い。実際、情報学部情報社会学科において講義やゼミで統計学を学ぶ者の半数以上は、高校時代に文系コースを履修している。

統計学は、現実社会をデータから読み取るという点で文系要素を含むが、それと同時に、分析する際に使う計算式を理解する点などでは理系要素も含んでいる。そのため一般的に、文系学生は統計学を学ぶときその理数的側面に躓きを感じるようになる。文工融合の教育を実践している静岡大学情報学部でも、理数系分野に躓きを感じる文系学生が多いものと思われる。しかし統計学が現実社会を理解する上で重要な武器である以上、文系学生といえども、統計学の学習をおろそかにすることはできない。従って、文系学生の統計学学習における躓きの実態とその要因を分析することは、文工融合の教育を実践する上で欠かせないことであろうと考える。

一方で、統計学が理系的要素を含む学問分野のため、文系学生は統計学に興味を持たないとは一概に言えない。実際、データから得た分析結果を読み取ること非常に興味を持ち統計学を学んでいる学生も少なからずいる。躓きを分析することと同時に、統計学のどのような点に興味を持っているかを分析することも躓きを軽減するために役立つのではないだろうか。

統計は、人文・社会科学や経済学、自然科学など文系・理系を問わず広い分野で必須である。統計学は、文系学生にとって最も身近な理数系分野の一つであろう。よって、統計学に対する苦手意識を調査することは静岡大学情報学部に限らず、一般に社会科学系の分野を学ぶ学生にとっても非常に有効なことだと考える。

以上のことから本研究では、大学生の現状に適応した統計学不安尺度を作成して大学生の統計学に対する意識を調べ、それと統計学の履修状況、特にその躓きとの関連について分析する。調査対象は文系学生であるが、比較対象として理系学生の統計学に対する態度も調べ、その違いについても考察する。

1. 2 先行研究の検討

特に社会科学を学ぶ学生における統計学に対する苦手意識に関して、欧米では、統計学不安尺度が構成され利用されているが、日本ではまだ不安尺度に関する研究は十分には進んでいない。ただし、山田他(2004)は英語版の Statistical Anxiety Rating Scale (Mji, & Onwuegbuzie, 2004) をもとに日本語版の統計学不安尺度を作成し、それを用いて統計学不安意識調査を行っている。調査対象者は、静岡大学工学部と情報学部の学生 138 名である。調査に用いられた統計学不安尺度は、基本的には英語版を翻訳した 40 項目である。以下、山田(2004)を参考にして、そこで指摘された問題点を整理するとともに、それを克服するための方策について検討する。

英語版の統計学不安尺度は、主として社会科学系の大学院に在籍する学生を対象として作成されたものである。これらの学生は、既に学部時代に統計学の学習を行っており、大学院ではそこで得た知識・技術を前提として研究を行うことが要求される。ところが、学生の多くは文系であるため統計学の学習が不十分であり、その結果として統計学が大学院での研究を進める上での大きな障害、そしてそれ故に不安材料となっている。英語版不安尺度はこのような背景で作成されたものであるため、まずは実態として「学生が統計学に関してどの程度の知識・技術を保持しているか」を把握し、その上で「統計学に対する抵抗感(不安感)」を測定するように構成されている。山田(2004)は 40 項目に対する因子分析の結果、次の 4 つの因子を抽出している。

- ①統計関係の受講経験
- ②統計処理の技能
- ③統計は難しいとの評価
- ④統計は不必要で嫌いだとする意見

得られた①、②の項目群は「学生が統計学に関してどの程度の知識・技術を保持しているか」を把握するため、そして③、④の項目群は「統計学に対する抵抗感(不安感)」を測定するための項目に対応している。

このような不安尺度を本研究で用いる場合、2つの問題点がある。まず、本研究で対象としているのは学部学生であり、高校数学での簡単な学習経験を除けば統計学の初学者である。従って、①、②に対応する項目は本研究では必要ない。次に「抵抗感・不安感」についてはその程度(強さ)を測定することに主眼が置かれているために、同じような内容が繰り返し問われる形式になっている。確かに、これは信頼性の高い不安尺度を構成するために必要なことではあるが、その不安の内容の多様性に迫ることはできない。実際、山田他(2004)の調査結果では、「統計学は必要であり嫌いではないが、難解である」との回答が多くを占めており、学生の統計学に対する態度は単に好き-嫌いの軸だけで捉えることはできない多様な側面を持っていることが明らかになっている。

このような学生の統計学に対する態度を把握するためには、単に「抵抗感・不安感」を

測定するのではなく、どのような側面に抵抗感を感じ、どのような側面に興味や必要性を感じているか等を調べる必要がある。そこで本研究では、③「難しい」、④「不必要・嫌い」の内容をさらに掘り下げ、より学生の現状に適した統計学不安尺度の構成を目指す。

また、山田他（2004）で使われた不安尺度は、先にも述べたとおり、欧米で作成された不安尺度を翻訳したものである。そのため、例えば、「中学校で習った数学さえ理解する事ができないのに、統計学を習得するなんてとても無理だ」「実生活では決して統計学を使わないだろうに、どうして統計学の全てのことを勉強しなければならないのかと不思議に思う」という項目に見られるように、「無理だ」や「決して～ない」のような強いニュアンスを持つ言葉が多い。しかし、日本人の場合には、強い否定や強い肯定を使うと、回答者が「そう思う」「そうは思わない」というところに回答しづらいと考え、本研究の不安尺度では可能な限りそれらの言葉を使うことは避けることとした。

1.3 新たな統計学不安尺度の枠組み

本研究では、山田他（2004）の調査で使用した統計学不安尺度をもとに、それを改善する形で不安尺度を作成した。先にも述べたように、山田他（2004）では、各項目が「統計関係の受講経験」「統計処理の技能」「統計は難しいとの評価」「統計は不必要で嫌いだとする意見」という4つの因子に分類されている。そこでそれらを参考に、本研究ではより具体的に、そして多様な観点から不安の実態に迫るため、6つの質問項目分野を設定した。そして、それぞれの分野を調査・測定するために必要な項目を検討するという手順で不安尺度を作成していった。以下が質問項目の6分野である。

分野1. 全体的な忌避感

分野2. 統計学と数学との関係

分野3. 統計学習における壁

分野4. 統計学を学ぶ楽しさ

分野5. 統計学の必要性

分野6. 一般的学习意欲

まず、学生が統計学を好むのか避けたいのかという全体的忌避感を尋ねる質問分野1を設定した。学生は、自分自身ではなぜ統計学が好きなのか、なぜ嫌いなのかを意識していないかもしれない。そこで、全体的忌避感を調査し、分析する際に「なぜ統計学が好き、または嫌いなのか」という形で展開する。

全体的忌避感と同時に、具体的にどのような点に躓きを感じるのかを調査する。山田（2004）の結果を見ると、「統計学は難しいとの評価」という因子に分類された項目に、「統計学がそれほど数学的でなかったら、統計学も楽しいと思う」や「統計学はそんなに嫌で

はないけれども、あまりに数学的すぎる」などがある。それらの問いに対して、文系学生が「そう思う」と回答する傾向が強く、統計学の数式を使う点に躓きを感じる学生が多いことが伺える。

そこで、今回使用する不安尺度の項目に、統計学と数学との関係を調べる項目が必要だと考え、分野2「統計学と数学との関係」を設定した。この分野では、学生は統計学をどの程度理数系科目であると感じているのかを調べるのが目的となる。

今回の調査票では、より学生の実態を反映した不安尺度を作るため、統計学を学習する文系学生に対してインタビュー調査を実施した。そこで得られた主要な意見である「統計学学習時に感じた躓き」を反映させ、尺度を作成した。その質問分野が分野3「統計学習における壁」である。またインタビューでは、「学習における躓き」を感じながらも、「統計学を学ぶことの楽しさ」に言及する意見も得られた。そこで、統計学に対する苦手意識だけではなく、どのような点に興味を持っているのかを調べることを目的に、分野4「統計を学ぶ楽しさ」という分野を設定した。「学ぶ楽しさ」は、学習の壁を乗り越える力になる可能性があり、不安を解消させるための方策を提供してくれる可能性もある。

また、先行研究の結果から「統計学は不必要で嫌いだ」という因子が抽出された。そこで、統計学に必要性を感じるかどうかを積極的な学習へと導く要因となると考え、分野5「統計学の必要性」を設定した。

また、学生の一般的な学習意欲の高低は、統計学に限らず大学での学習に影響を与えていると考え、「一般的学習意欲」を調査する質問分野6を加えた。

1. 4 分野別項目の詳細

(1) 全体的忌避感

学生は、統計学を避けたいという意識なのだろうか、統計学を積極的に学びたいという意識なのだろうか。以下が、学生が統計学に対してどれ程の忌避感を感じているかを調査するための項目である。

1. うまく理由は言えないが、とにかく統計学は好きではない
2. 統計学には人間味を感じない
3. 統計学を習得できたらよいと思う
4. 統計学が必修でなければよいと思
5. 多くの数値や図表が含まれている新聞や雑誌記事を読むのは抵抗がある
6. 統計学的分析を含む学術論文や本を読んでも嫌にならない
7. 統計データに頼りすぎるのは問題だと思う

1は、統計学が好きか嫌いかという大きな枠で捉える意識が統計学の理解に影響を与え

ているのではないかと考え作った質問である。忌避感を感じる要因として、データ列ばかりを扱う統計学に人間味を感じないのではないかと、そもそも統計学を習得したいと思っているのかどうかという考えから2、3、4を加えた。また、統計学の講義という特別な環境ではなく、普段の生活で触れる統計に対しての意識を調査するため、5、6、7の質問を加えた。

(2) 統計学と数学との関係

統計学を学習する際、避けて通れないのが数式理解や表グラフ読み取りである。これこそが統計学が理数系科目であるという所以である。学生は、統計学と数学とは理解するに当たり密接な関係があると思っているのだろうか。以下の項目を用意した。

1. 統計学がそれほど数学的でなかったら、統計学も楽しいと思う
2. 一般的に、数学が理解できないと統計学を理解することは難しいと思う
3. 自分は数学が好きである
4. 数値を扱うのは好きではない
5. 数値には冷たい印象がある
6. 様々な計算式を使って結果を出すことを面白いと感じる

これらの質問項目に対する回答と、分野1の「全体的忌避感」に関する質問項目に対する回答との関係を分析し、学生が数学と統計学の関係をどう捉えているのかを考える。

(3) 統計学における壁

統計学を学習するにあたり、学生は様々な壁にぶつかる。これが本研究で最も注目すべき点である。学生はどのようなときに躓きを感じるのであろうか。

統計学は、大きく分けて2つの側面を持つ。1つ目はデータを収集し分析するという面、もう1つは、分析した結果を読み取るという面である。前者は、分析する過程に数式を使うため理数的側面である。後者は、分析結果を現実社会とリンクさせる作業であり、文系的側面である。統計学を学ぶ際、文系学生の多くが躓くのは統計学の持つ理数的側面であろう。そこで以下の質問項目を用意した。

1. 統計学には難しい計算式ばかりが使われる
2. 統計処理を行う際、多くの数値データを扱うのに抵抗を感じる
3. 図や表から傾向を読み取ることが得意である

先にも述べたように、項目作成にあたり文系学生にインタビューをした。その際、統計学

の理数的側面として「数式を扱うこと」、「大量の数字データを扱うこと」、「図や表を扱うこと」への苦手意識があるという意見が多かった。そこで、上記1の計算式に関する質問と2、3の数値に関する質問を取り入れた。

学生が統計学に躓くのは、理数的側面だけではない。学生は、大学1年次に統計学を履修するわけだが、初めて耳にする専門用語や概念が多いはずである。そして、そのような点にも学生は躓くと考えられる。その過程を反映させた質問が以下の4、5である。

4. 統計学の講義では、専門用語が多用されるので混乱する事がある
5. 統計学には覚えなければならない概念・事柄が多すぎる

理数的側面や専門用語の他に、以下の質問を用意した。

6. 統計学で学んだ事が、実際にどのようなことに使われているかわからない
7. 自分はもともと勘に頼る性格なので、統計学のもつ論理的な面に苦手意識を感じる

今学習していることがいったい何に使われているのか、目的は何なのかがわからなければ、主体的に学習に取り組むことは難しいであろう。学生は統計学がどのような事に使われているのか具体的にわからないため、学習に取り組む姿勢が消極的になってしまうのではないだろうか。そこで、質問6を用意した。また、どのようなことに使われているかわからないことから、統計学に価値を感じない可能性があるため、質問7を作成した。

(4) 統計学を学ぶ楽しさ

学生は統計学のどのような点に興味を持っているのだろうか。統計を行う過程を考えると、大きく分けて、まず数式から結果を出す、次に結果を読み取る、結果を現実社会と照らし合わせるという過程がある。それぞれを質問項目としたのが以下である。

1. データの分析をすることを興味深いと思う
2. データ解析によってどのような結果が得られるかワクワクする
3. 新聞・雑誌等に掲載紹介されている統計のデータを読むのは面白い

興味を持っている点を分析し、今後の教授法に活かしていくことができれば、学生の持つ統計学不安を少しでも改善できるのではないであろうかと考え、上記の質問を設定した。

(5) 統計学の必要性

統計学は、先にも述べたが、文系でも多くの学生が必要とする学問のはずである。この点、学生はどの程度強くその必要性を感じているのだろうか。統計学を自分自身にとって必要と感じているのか、不必要と感じているのか、また一般的に必要と感じるか、不必要と感じるかを調査する。

1. データに基づかない議論は信用できない
2. 誰でも、多かれ少なかれ統計解析の知識・技能を必要とすると思う
3. なぜ自分の専門分野に統計学が必要なのかわからない
4. 将来、統計学の知識を積極的に役立てていきたい
5. 統計学には多くの時間・労力を傾けるだけの価値はない
6. 現実には起こっている事柄を正しく把握するには、統計学が必要だと思う

必要だと感じて講義を受講する学生と不必要と感じて講義を受講する学生とでは、理解に差があるのだろうか。これらの質問と、統計学の講義で毎回採っているアンケートの回答との関連を調べる。また、なぜ必要と感じるのか、なぜ必要と感じないのかという点に関しても、これらの質問項目に対する回答から分析する。

(6) 一般的学习意欲

統計学に限らず、学生の学習意欲を調査する。以下がその質問項目である。

1. テストや課題があれば勉強するが、なければ勉強しないことが多い
2. 物事を学ぶ上で、基礎は大切だと思う
3. 学習中、わからない事があればわかるように努力する
4. 予習・復習をする習慣がある
5. 大学の講義には深い専門性を求めている

質問1, 2, 3, 4により学生の学習状況の実態が見える。またこれらの項目から、学生が求めている知識像が浮かび上がる可能性もある。

1. 5 アンケート票

以下に作成した「統計学不安尺度」36項目を示す。これらの項目は4カテゴリの評定尺度でその回答を求めた。

Ⅱ 統計学に対する意識に関する質問

(最もあてはまる数字に○をつけてください)

	4 そう思う	3 やや 思う	2 あまり 思わない	1 そうは 思わない
1. 自分は数学が好きである	4	3	2	1
2. 統計学を習得できたらよいと思う	4	3	2	1
3. 統計学で学んだ事が、実際にどのような ことに使われているかわからない	4	3	2	1
4. 新聞・雑誌等に掲載紹介されている統計の データを読むのは面白い	4	3	2	1
5. うまく理由は言えないが、とにかく 統計学は好きではない	4	3	2	1
6. 現実には起こっている事柄を正しく把握するには、 統計学が必要だと思う	4	3	2	1
7. データに基づかない議論は信用できない	4	3	2	1
8. 多くの数値や図表が含まれている新聞や 雑誌記事を読むのは抵抗がある	4	3	2	1
9. 統計学がそれほど数学的でなかったら、 統計学も楽しいと思う	4	3	2	1
10. 様々な計算式を使って結果を出すことを 面白いと感じる	4	3	2	1
11. 将来、統計学の知識を積極的に役立てたい	4	3	2	1
12. 統計学には覚えなければならない概念や 事柄が多すぎる	4	3	2	1
13. 誰でも、多かれ少なかれ統計解析の知識・ 技能を必要とすると思う	4	3	2	1
14. 統計学的分析を含む学術論文や本を 読んでも嫌にならない	4	3	2	1
15. 統計データに頼りすぎるのは問題だと思う	4	3	2	1
16. テストや課題があれば勉強するが、なければ 勉強しないことが多い	4	3	2	1
17. 一般的に、数学が理解できないと統計学を 理解することは難しいと思う	4	3	2	1

- | | 思う
4 | やや
思う
3 | あまり
思わない
2 | 思わない
1 |
|---|---------|---------------|------------------|-----------|
| 18. 数値を扱うのは好きではない | 4 | 3 | 2 | 1 |
| 19. データの分析をすることを興味深いと思う | 4 | 3 | 2 | 1 |
| 20. 統計学の講義では、専門用語が多用されるので混乱する事が多い | 4 | 3 | 2 | 1 |
| 21. 物事を学ぶ上で、基礎は大切だと思う | 4 | 3 | 2 | 1 |
| 22. 統計処理を行う際、多くの数値データを扱うことに抵抗を感じる | 4 | 3 | 2 | 1 |
| 23. 学習中、わからない事があればわかるように努力する | 4 | 3 | 2 | 1 |
| 24. 統計学が必修でなければ良いと思う | 4 | 3 | 2 | 1 |
| 25. 図や表から傾向を読み取ることが得意である | 4 | 3 | 2 | 1 |
| 26. データ解析によってどのような結果が得られるかワクワクする | 4 | 3 | 2 | 1 |
| 27. 統計学には人間味を感じない | 4 | 3 | 2 | 1 |
| 28. 統計学では難しい計算式ばかりが使われる | 4 | 3 | 2 | 1 |
| 29. 自分はもともと勘に頼る性格なので、統計学の持つ論理的な面に苦手意識を感じる | 4 | 3 | 2 | 1 |
| 30. 予習・復習をする習慣がある | 4 | 3 | 2 | 1 |
| 31. 数値には冷たい印象がある | 4 | 3 | 2 | 1 |
| 32. 統計学には多くの時間・労力を傾けるだけの価値はない | 4 | 3 | 2 | 1 |
| 33. 大学の講義には深い専門性を求めている | 4 | 3 | 2 | 1 |
| 34. なぜ自分の専門分野に統計学が必要なかわからない | 4 | 3 | 2 | 1 |

*アンケートは以上で終了です。アンケートへご協力ありがとうございました。

第2章 調査の実施と解析

2.1 調査の実施

<調査1>

日時：2005年11月中

場所：静岡大学情報学部 of 統計学に関する講義中

対象：静岡大学情報学部1年生161名（情報科学科76名 情報社会学科85名）

<調査2>

日時：2006年12月中

場所：静岡大学情報学部 of 統計学に関する講義中

対象：静岡大学情報学部1年生132名（情報科学科50名 情報社会学科80名）

静岡大学情報学部2・3年生70名（情報科学科22名 情報社会学科48名）

2.2 分析結果

2005年度実施の調査結果については、加藤(2006)で詳しく議論されているので、ここでは2006年度調査の解析結果について述べる。

(1) 単純集計

表1-1は、29個の不安尺度項目、5個の一般的学習態度に対する単純集計の結果である。回答は、1～4までの評定カテゴリで求めたが、ここでは「思う」、「そう思う」を肯定カテゴリとしてそれらへの回答比率（肯定率）、と回答カテゴリを数値とした場合の平均値と標準偏差を示した。表は、肯定率の高い順に並び替えて示した。また、右端のグラフは、回答率を棒グラフとして示したものであり、真ん中の線は50%の位置を示す。

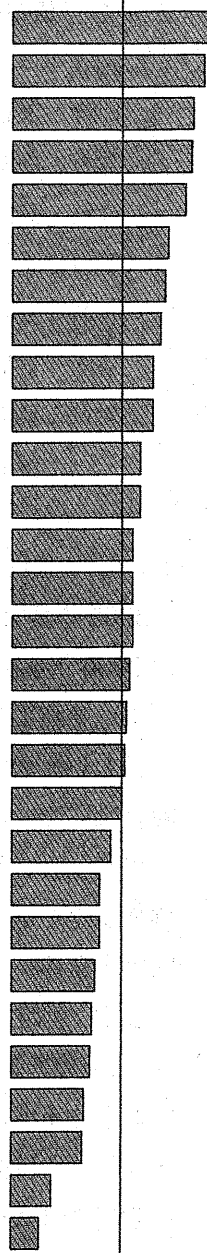
まず、肯定率の高い範囲には、「統計学が必要」、「統計を習得したい」、「データ分析が興味深い」等の統計学学習に対する積極的な評価がある一方で、「専門用語が多用されている」、「概念や事柄が多い」等、統計の学習における困難さも多く支持を得ている。つまり、全体的に、統計の学習に対する興味・関心は高くその必要性も感じているが、同時に、その学習には大きな困難が伴うとの認識である。これらには8割近い、あるいはそれ以上の回答が集まっており、これが統計学に対する一般的な評価と言える。

統計学の数学的側面に対する評価、例えば「数式を使って結果を出すことが面白い」、逆に「数学的でなければ楽しい」、「多くの数値データを扱うことに抵抗感がある」等はいずれも50%台であり、その評価は分かれている。調査対象には理系の情報科学科も含まれており、文系の情報社会学科との間で意見が分かるためであろう。「数学が好き」の肯定率は約5割である。さらに図表の扱いに関して「図や表から傾向を読み取ることが得意」は4割を下回っており、数値・数式だけでなく、図表も統計を学ぶ上での障害になっているようである。

表 1-1 単純集計

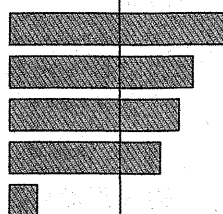
NO.	統計不安尺度	肯定率	平均	標準偏差
6	現実起こっている事柄を正しく把握するには、統計学が必要だと思う	90.1	3.2	0.413
20	統計学の講義では、専門用語が多用されるので混乱する事が多い	87.2	3.3	0.298
2	統計学を習得できたらよいと思う	82.3	3.1	0.510
12	統計学には覚えなければならない概念や事柄が多すぎる	81.3	3.1	0.314
19	データの分析をすることを興味深いと思う	78.3	3.0	0.319
28	統計学では難しい計算式ばかり使われる	70.9	2.9	0.483
13	誰でも、多かれ少なかれ統計解析の知識・技能を必要とすると思う	69.5	2.8	0.372
17	一般的に数学が理解できないと統計学を理解することは難しいと思う	67.5	2.9	0.466
7	データに基づかない議論は信用できない	64.0	2.8	0.544
15	統計データに頼りすぎるのは問題だと思う	64.0	2.7	0.488
4	新聞・雑誌等に掲載紹介されている統計のデータを読むのは面白い	58.6	2.7	0.500
24	統計学が必修でなければ良いと思う	58.1	2.7	0.502
9	統計学が数学的でなかったら統計学も楽しいと思う	55.2	2.7	0.640
10	様々な数式を使って結果を出すことを面白いと感じる	54.7	2.6	0.651
11	将来、統計学の知識を積極的に役立てたい	54.7	2.6	0.541
22	統計処理を行う際、多くの数値データを扱うことに抵抗を感じる	53.7	2.6	0.564
1	自分は数学が好きである	52.2	2.5	1.010
3	統計学を学んだ事が、実際にどのようなことに使われているかわからない	51.7	2.5	0.717
26	データ解析によってどのような結果が得られるかわくわくする	50.2	2.6	0.620
5	うまく理由は言えないがとにかく統計学は好きではない	45.3	2.5	0.613
18	数値を扱うのは好きではない	40.4	2.4	0.669
27	統計学には人間味を感じない	39.9	2.4	0.628
31	数値には冷たい印象がある	38.4	2.3	0.718
25	図や表から傾向を読み取ることが得意である	36.5	2.4	0.499
29	自分はもともと勘に頼る性格なので、統計学の持つ論理的な面に苦手意識を感じる	36.0	2.3	0.709
8	多くの数値や図表が含まれている新聞や雑誌記事を読むのは抵抗がある	33.5	2.3	0.758
34	なぜ自分の専門分野に統計学が必要なのかかわからない	32.5	2.2	0.778
14	統計学的分析を含む学術論文や本を読んでも嫌にならない	18.7	2.0	0.678
32	統計学には多くの時間・労力を傾けるだけの価値はない	12.8	1.9	0.685

0 100(%)



No.	一般的学習態度	肯定率	平均	標準偏差
21	物事を学ぶ上で基礎は大切だと思う	98.5	3.7	0.422
16	テストや課題があれば勉強するが、なければ勉強しないことが多い	83.3	3.2	0.557
23	学習中、わからないことがあればわかるように努力する	77.3	2.9	0.307
33	大学の講義には深い専門性を求めている	69.0	2.9	0.357
30	予習・復習をする習慣がある	12.8	1.8	0.747

0 100(%)



統計に対する強い忌避感、例えば「統計学には人間味を感じない」、「数値には冷たい印象がある」等に対する肯定率は4割程度である。無視できない大きさで、統計を嫌う集団が存在している。

「多くの数値や図表が含まれる新聞記事などを読むのに抵抗がある」とする者は3割程度、つまり7割は抵抗が無いとしているのに対し、その内容が高度になり「統計学的分析を含む学術論文、本を読んでも嫌にはならない」となった場合には、その肯定率は2割を下回る。それほど高度な内容を期待しているわけではない。

なお、「統計学には時間をかけて学ぶ価値はない」に対する肯定率は1割程度であり、先にも述べたように、多くの者は「統計は必要であるし、それほど高度な内容でなければ、しっかり身につけておきたい分野である」と考えている。

(2) 項目間の関連

項目間の関連を整理するために、統計学不安尺度29項目に対して因子分析を適用した。表1・2は、得られた第5因子までの解にバリマック回転を適用した結果で、それぞれの項目に対する因子負荷量を表としてまとめたものである。また、各因子に関して高い因子負荷量を示す項目には網掛けをした。基本的に各項目は1つの因子に対応させる形で整理したが、項目によっては複数の因子に高い負荷を示す場合もある。なお、第5因子までの寄与率は40%であり、必ずしも寄与率は高くない。以下では各因子の内容を吟味する。

<第1因子>

負荷量の高い項目は、「数学が好き」、「数式を使うのが面白い」、「図表を読むのが得意」等の数学学力の高さを表すものであり、さらに、「データ解析にワクワクし」、「分析結果に興味を持ち」、そして「統計データの解析に面白さを感じる」等、統計に対して強い興味・関心を示す項目も含まれている。つまり、第1因子で高い因子スコアを持つ者は、高い理系的学力によって統計を理解し、その理解ゆえに統計に対しても興味・関心を抱いている者、ということになる。従って、第1因子は「理系的学力に裏付けされた統計に対する興味・関心」と解釈できる。

<第2因子>

第2因子で正の負荷量を示したのは、「将来、役立てたい」、「統計を習得したい」、「統計解析の知識、技能が必要」である。一方、負の負荷量の項目は「統計学が必修なのはイヤだ」、「学習する価値はない」、「統計の必要性が分からない」とするもので、正の項目とは統計の必要性、価値に対する認識が全くの反対で、それ故学習意欲も低いとする項目群である。つまり、第2因子は「統計の必要性に対する認識」に対応しており、正のスコアは必要性を強く感じていることを意味する。

＜第3因子＞

ここに含まれた項目は、いずれも統計に対する忌避感を表している。例えば、「論理的な思考が苦手」で、従って「統計には人間味を感じない」。さらに「統計学がどのように利用されているか分からない」し、もともと「図表や多くの数値には抵抗感がある」。つまり、「統計は嫌い」である。つまり、第3因子は「統計に対する忌避感」に対応しており、正のスコアは忌避感をより強く感じていることを意味する。

＜第4因子＞

第4因子は3項目で、いずれも統計の学習に対する壁として認識されている特性である。つまり、統計は「専門用語が多く」、「概念や事柄も多く」、さらに「難しい計算式を使う」ために、その学習には困難を感じる。つまり、第4因子は「統計学習の壁」に対応してお

表 1-2 因子負荷量

変数名	第1因子	第2因子	第3因子	第4因子	第5因子
数式面白い	0.605	0.154	0.027	-0.180	-0.213
データ解析ワクワク	0.581	0.268	-0.244	0.058	-0.020
図表得意	0.559	0.009	-0.133	-0.110	-0.009
数学好き	0.508	0.001	-0.124	-0.182	-0.405
分析興味深い	0.490	0.40	-0.166	-0.010	0.022
統計データ面白い	0.445	0.302	-0.132	0.049	0.023
知識を役立てる	0.306	0.671	-0.015	-0.152	-0.016
統計学習得希望	0.399	0.567	0.073	-0.237	-0.066
統計解析の知識・技能必要	0.039	0.502	-0.047	0.032	0.114
統計学必修イヤ	-0.155	-0.416	0.260	0.206	0.052
価値はない	-0.024	-0.509	0.206	0.167	0.036
統計学必要性わからない	-0.095	-0.570	0.319	0.131	0.077
統計学利用法不明	-0.162	-0.052	0.547	0.149	-0.116
人間味感じない	0.088	-0.374	0.527	0.122	0.076
数値図表抵抗	-0.246	-0.082	0.496	0.101	0.042
論理的苦手	-0.079	-0.242	0.485	0.131	0.306
統計学嫌い	-0.415	-0.367	0.449	0.086	0.130
数値データ抵抗感	-0.344	-0.068	0.447	0.181	0.204
統計学的学術論嫌でない	0.280	0.205	-0.345	-0.304	-0.087
専門用語多用	-0.126	-0.117	0.149	0.682	0.014
概念や事柄が多すぎる	-0.180	-0.126	0.165	0.677	-0.063
難しい計算式	-0.047	-0.160	0.118	0.613	0.202
数値嫌い	-0.466	-0.229	0.307	0.075	0.526
統計学必要	0.064	0.330	-0.156	0.063	0.480
非数学的統計学楽しい	-0.051	0.032	0.233	0.364	0.435
数値冷たい	0.044	-0.213	0.30	0.148	0.416
データ必要	0.043	0.006	-0.071	-0.032	0.371
統計学数学必要	-0.105	0.003	0.127	0.001	0.366
統計データ頼り問題	0.235	-0.006	-0.027	-0.084	0.094

り、正のスコアはその壁をより強く認識していることを意味する。表 1-1 では、これら 3 項目がいずれも 8 割以上の肯定率を示していたことを見たが、この「壁」は多くの学生に共通して認識されている。

<第 5 因子>

第 5 因子には必ずしも方向性が一致しない複数の項目が含まれる。つまり、「数値は嫌い」であり、「数値には冷たい印象」はあるが、「統計学は必要であり」、「データで議論することも必要」であるとも感じており、従って「統計学が数学的でなければ楽しい」と考えている。つまり、統計に対する必要性は認識しているものの、統計の数学的側面に大きな壁を感じている状態である。その意味で、第 5 因子は、「統計の数学的側面による躓き」であり、特に文系の学生の感ずる「壁」に対応していると解釈できる。

(3) 学科・コースと統計学不安尺度との関係

静岡大学情報学部には理科系の情報科学科と文系の情報社会学科より構成されている。これらの学科、学習の内容はもちろんのこと、入試制度まで全く異なっており、前者は数学を個別試験にも課す理科系入試、後者は小論文等を個別試験に課す文系入試を行っている。それに加え、情報学部では 2004 年度から、以下の 3 つの教育プログラムから構成される新しい教育カリキュラムをスタートさせた。

CS (Computer Science Program : 計算機科学プログラム)

情報技術の開発・設計

IS (Information System Program : 情報システムプログラム)

実社会の中で情報技術を活かすための情報システムの開発・運用

ID (Information Society Design Program : 情報社会デザインプログラム)

情報技術・情報システムにもとづく新たな情報社会形成の考察・設計

この制度は 2 学科の学生を、上記の 3 つの教育プログラムのいずれかで教育するもので、この中で IS プログラムは、情報科学科、情報社会学科の両学科からの受講を可能としている。つまり、情報学部の学生は 2 つの学科に分かれるだけでなく、学科とプログラムを組み合わせた以下の 4 つの群に分かれて学習する。

①情報科学科 CS プログラム専攻

②情報科学科 IS プログラム専攻

③情報社会学科 IS プログラム専攻

④情報社会学科 ID プログラム専攻

これらの学生群は、学科特性、及びプログラム特性に関連して、異なった学力特性を持つと考えられる。それは、学生の理系－文系の学力特性で、上の①～④は、最も理系的な情報科学 CS から最も文系的な情報社会学科 ID まで、文系色が強まる系列として並べた。このように異なった学力特性を持つ学生は、統計学不安尺度に関してどのような違いがあ

るのか、その関連を以下で検討する。

表 1-3 は、上記の 4 群ごとに第 1 因子～第 5 因子までの因子スコアの平均と標準偏差を表としてまとめたものであり、図 1-1 は、平均値をレーダーチャートで示したものである。

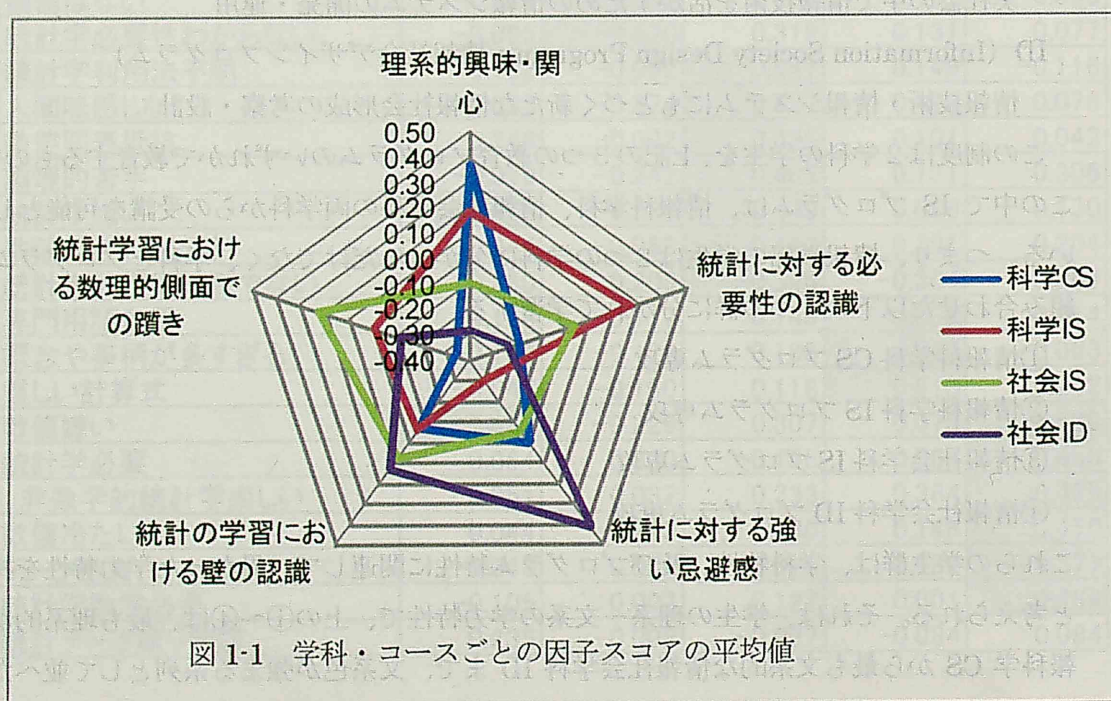
この図表に基づいて、各学生群の特徴をまとめる。

<情報科学科CSプログラム>

「統計の理系的側面」に対しては最も高い興味・関心を寄せており、理系的学力の高さのためにその学習においても「数理的側面での躓き」を感じることもし少ない。ただし、「統計の必要性に対する認識」は低く、その活用に関してはあまり興味を持っていない。

表 1-3 学科・コースごとの因子スコアの分布（平均と標準偏差）

統計量	因子	学科・コース			
		科学CS	科学IS	社会IS	社会ID
平均値	理系的興味・関心	0.37	0.18	-0.10	-0.29
	統計に対する必要性の認識	-0.20	0.27	0.04	-0.24
	統計に対する強い忌避感	0.00	-0.29	-0.05	0.42
	統計の学習における壁の認識	-0.07	-0.05	0.08	0.14
	統計学習における数理的側面での躓き	-0.34	-0.01	0.23	-0.11
標準偏差	理系的興味・関心	0.76	0.73	0.90	0.95
	統計に対する必要性の認識	0.73	0.68	0.96	0.92
	統計に対する強い忌避感	0.74	0.81	0.84	0.79
	統計の学習における壁の認識	0.77	0.83	0.82	0.84
	統計学習における数理的側面での躓き	0.56	0.77	0.91	0.80
データ数		24	44	73	43



<情報科学科 I S プログラム>

「統計の必要性」に関する強い認識があり、従って「忌避感」も最も弱い。しかも、「理数的興味・関心」も情報科学科 CS に続いて高いために、「数理的な側面での躓き」も少ない。学ぶ意欲、学力の両方の面で、最も適しているといえる。

<情報社会学科 IS プログラム>

「統計の必要性」に関する認識は、情報科学科 IS に続いて高いが、「理数的興味・関心」が低いために「数理的な側面での躓き」を最も強く感じている。意欲が高いだけに、学習の困難さをより強く感じている。統計学不安尺度の意味では、最も不安の高い集団である可能性がある。

<情報社会学科 ID プログラム>

「忌避感」が極めて強い集団である。その背景にあるのは「理数的興味・関心」の低さであるが、それが原因で「必要性の認識」も低い。学習の困難さゆえに学習意欲を失っており、従って、「数理的な側面での躓き」もあまり強くは感じていない。学習の放棄の段階である。

(4) 統計学習に対する意識に基づく授業改善の方向

前項では、統計学不安尺度を用いることにより、4つの学生群が統計学習に関してそれぞれ異なった意識を持ち、異なった問題を抱えていることが明らかとなった。以下では、得られた結果に基づいて、学科・コース別に授業改善の方向について簡単に述べる。

<情報科学科 C S プログラム>

統計を学習する十分な理系的学力を有するが、統計をあまり役立つものとは考えていない者が多い。統計解析が社会的な事象に限らず様々な場面で有効であることを例によって示せば、彼らの興味を引きつけ、学習意欲を高めることがきるかもしれない。また、この群の学生は、卒業研究で実験を行うことが多く、その場合、そこで取得したデータの効果的な解析法、たとえば仮説検定等の手法を実験データ解析の場面で解説すれば、より効果的であると考ええる。

<情報科学科 I S プログラム>

統計を学習したいという意欲と、その学習を可能とするある程度の理系的学力も有している者が多い。彼らの意欲を低下させないためには、具体的なデータ解析例とともに、解析手法の理論的背景をキチンと説明する必要がある。もちろん、あまり難解なものは避けるべきであるが、データ処理をブラックボックス化せず、その仕組みを説明することによって、さらに興味・関心を高めることができる。

<情報社会学科 IS プログラム>

興味・関心は高いが、理数的学力が十分ではないために学習に困難を感じている者が多

い。従って、データ解析の具体例に関しては興味を示すが、理論的な解説には抵抗がある。理論的な解説については、できるだけ数式を用いず、その数式の意味することを例示するのが良い。その例示方法としては、本研究で開発した「統計を学習するためのシミュレーション・システム」(本報告書第2部)が有効と考える。

<情報社会学科 ID プログラム>

統計の学習を全く受け付けられない者も含まれる。本来、社会的な事象に関しては強い興味を抱いている学生群であり、その興味を高めるような具体性の高い解析事例を中心に講義を進める必要がある。内容もあまり深入りせず、簡単なデータ解析で得られる豊富な情報を例示することで、データ解析の面白さを伝えることが必要である。

2. 3 まとめ

従来の統計学不安尺度では、「抵抗感・不安感」についてその程度(強さ)を測定することに主眼が置かれているために、同じような内容が繰り返し問われる形式になっている。これは信頼性の高い不安尺度を構成するために必要なことではあるが、その不安の内容の多様性に迫ることはできない。その点、本研究で作成した不安尺度はその程度ではなく、多様性を明らかにすることを目的としており、前項の議論で十分にその機能が有効であることが確認できた。今後の改良の余地はあるが、統計学習に対する意識を測定するための適当な尺度であると考ええる。

<参考文献>

加藤絢子 (2006)、「文系学生における統計学に対する不安意識の解析」、静岡大学情報学部卒業研究、2006.3

山田文康・漁田武雄・高橋晃・新谷誠(2005)、「文系学生における理数系科目不安の実態把握とその解消を促進する教授法の開発」、情報学研究推進プロジェクト研究成果報告会、静岡大学、2005.4

Mji,A. , Onwuegbuzie, A.J. (2004), Evidence of Score Reliability and Validity of the Statistical Anxiety Rating Scale Among Technikon Students in South Africa, *Measurement and Evaluation in Counseling and Development*, Vol. 36, 238-251

Onwuegbuzie, A.J., (2004), Academic procrastination and statistics anxiety, *Assesment and Evaluation in Higher Education*, Vol.29, No.1, 3-19

Baloglu, M. (2003), Individual difference in statistics anxiety among college students , *Personality and Individual Difference*, 34, 855-865

第 2 部

統計的推測を学習するための
標本抽出実験
シミュレーション・システム

第1章 システムの概要

1.1 目的

統計的仮説検定や区間推定は、統計を学ぶ人達、特に文系の学生にとって大きな壁になっている。実際に仮説検定を使っている人でも、それを十分に理解して行っているわけではなく、単に示された手順に従って処理しているに過ぎないというケースも数多く見受けられる。

その大きな原因は、標本統計量の分布にあるものと思われる。たとえば、我々が調査データから平均値を求めたとすると、その平均値は、ある確率分布に従う標本平均の分布から得られた1つの実現値だということになる。しかし、我々が手にしているのは1つの標本平均値だけであり、これがある確率分布から得られた1つの値だと言われてもピンと来ないのは当然である。しかし、母集団からの標本抽出を何度も何度も繰り返し行うことができれば、複数の標本平均値が得られますので、その分布を考えることができるはずである。

この標本抽出実験をパソコン上の仮想的な母集団に基づいて行い、標本統計量の分布を理論の世界から現実の世界に引き出そうというのが、このシミュレーション・システムの目的である。標本統計量としては、平均値、比率、分散、相関係数等が対象で、その分布がどのような形状を示すのか、標本数を変化させるとその分布がどのように変化するか等を検討することができるよう設計した。また、簡単な仮説検定、区間推定の方法については、その解を与えるツールを提供する。

このようなシミュレーション・システムは、既に、山田(2000)、山田・** (1999)で開発されているが、実際の講義での利用を通して様々な問題点が指摘され、さらにより理解しやすいシステムにするための改良点も明らかになってきた。そこで、本研究では、特に以下の点に留意してシステムの再構築をおこなった。

①母集団分布の多様性を反映できること

母集団分布を一様分布に限定していたものを、以下の分布に拡張した

* 正規分布

* 正に歪んだ分布

* 双峰性分布

* 不定型分布

(正に歪んだ分布、双峰性分布、不定型分布は乱数を用いて発生させるもので、利用者はシステムが提供する様々な分布の中で、適当なものを選択することができる)

②理論分布との対応を明確にすること

シミュレーションによって得られた分布と理論分布の対応関係を理解しやすい形で提示する。

③推定・検定のツールについては、可能な限り利用者のデータ処理プロセスを組み込み、その処理プロセスを通して、推定・検定の基本的な考え方が理解できるようにすること

④ヘルプ機能を追加すること

従来のシステムにはヘルプ機能がなかったため、受講者の自習用には必ずしも適していない。自習用も十分な学習効果を上げるため、全てのシミュレーション処理画面に、処理の内容、処理の手順を表示するためのヘルプボタンを設定する。

⑤具体的な標本抽出の過程を経験できること

仮想的な標本抽出だけでは、具体的な標本抽出過程を経験することができない。シミュレーション・システムでの学習を確実なものとするためには、具体的に母集団を設定し、それからの抽出過程を示す必要がある。

⑥分散分析を追加すること

従来の、平均値、比率、2変量間の関連についての推定・検定に、一要因の分散分析、二要因の分散分析を加える。

1. 2 システムの構成

本システムは、①標本抽出実験サブシステム、②統計的検定・推定サブシステムの2つのサブシステムから構成する。

(1) 標本抽出実験サブシステム

パソコン上に仮想的に設定された母集団から無作為抽出を行い、得られた標本データに基づいて平均、比率、分散等の標本統計量の分布を具体的に示すためのシステムである。表 2-1 は、このサブシステムで扱う標本抽出の場面設定を示したもので、この設定場面ごとに標本抽出のシミュレーションを行うことができるように構成されている。

表 2-1. 本システムにおける標本抽出実験の場面設定

場面設定		内容
無作為抽出	母集団からの標本抽出	母集団から無作為抽出した標本データの特性を調
1標本	標本平均の分布	1つの母集団から得た無作為標本に基づく標本平均がどのような分布をするか調べる
	標本比率の分布	1つの母集団から得た無作為標本に基づく標本比率がどのような分布をするか調べる
	標本分散の分布	1つの母集団から得た無作為標本に基づく標本分散がどのような分布をするか調べる
	不偏分散の分布	1つの母集団から得た無作為標本に基づく不偏分散がどのような分布をするか調べる
	標本平均を標準化した場合の分布	1つの母集団から得た無作為標本に基づく標本平均を不偏分散を用いて標準化した場合に、標準化平均がどのような分布をするか調べる
2標本	標本平均値の差の分布	2つの母集団から独立に得た2つの無作為標本に基づく標本平均の差が、どのような分布となるか調べ
	標本比率の差の分布	2つの母集団から独立に得た2つの無作為標本に基づく標本比率の差が、どのような分布となるか調べ
3標本以上	3つ以上の平均値	全く同一の3つ以上の母集団から得た無作為標本に基づいて求めた相関比、F値がどのような分布を示すか調べる
	3つ以上の比率	3つ以上のカテゴリに対する母比率が与えられている場合、その母集団から無作為に抽出したデータに基づく χ^2 統計量の分布を調べる
2変量間の関連	相関係数	母相関係数が0の母集団から抽出した無作為標本に基づく標本相関係数が、どのような分布を示すか
	クロス表の関連度	全く無関連な2つの質的変数によって構成される母集団から得た無作為標本に基づいて求めた χ^2 統計量の分布を調べる
理論分布	正規分布	平均、分散の値による正規分布の形状の変化について調べる
	t分布	自由度による分布の形状の変化について調べる
	χ^2 分布	自由度による分布の形状の変化について調べる
	F分布	自由度(2つ)による分布の形状の変化について調

(2) 統計的検定・推定サブシステム

いくつかの典型的な場面で、区間推定、仮説検定を行うためのサブシステムであり、具体的には表 2-2 に示す推定・検定を行うことができる。

- ①分析対象データ及び推定・検定のための条件（帰無仮説、有意水準等）を設定する
- ②帰無仮説が正しい場合の標本統計量の分布を表示する
- ③その分布の中で、標本データから得られた実現値の位置を表示する
- ④標本統計量の分布における実現値の位置から、結論を導く

表 2-2. 本システムで利用できる推定・検定の方法

種別	推定・検定の内容	備考
区間推定	母平均の区間推定	母分散が未知／既知
	母比率の区間推定	
母平均の検定	1 標本の平均値の検定	母分散が未知／既知
	2 標本の平均値の差の検定	母分散が未知／既知
母比率の検定	1 標本の比率の検定	
	2 標本の比率の差の検定	
	χ^2 適合度検定	
2 変数間の関連度の検定	相関係数の検定	
	クロス表の独立性の検定	
分散分析	一要因	被験者内／被験者間
	二要因	被験者内／被験者間

1. 3 システムの導入

(1) 動作環境

本システムは、Microsoft Windows XP の元で動作するシステムとして、Microsoft Excel 2003 のマクロ言語 VBA (Visual Basic for Applications) を用いて作成した。従って、システムの運用は全て Excel の元で行う。

(2) ソフトウェアのフォルダ構成

シミュレーション・システムは以下のフォルダ構成で保存してある。ヘルプファイル等も全て含まれているので、全ファイルを一括してダウンロードする必要がある。

表 2-3 にシステムのファイル構成を示す

表 2-3. 本システムのファイル構成

フォルダ	フォルダ	フォルダ/ファイル	ファイル	ファイル名	
統計的仮説検定・推定	インストール クション	カイ2乗検定	2	手順1.doc	手順2.doc 等
		一要因分散分析	13	上限値(一要因).txt	効果の推定(等
		一要因分散分析(被験者内)	21	上限値(二要因).txt	主効果による等
		二要因分散分析	21	上限値(二要因).txt	主効果による等
		二要因分散分析(被験者間内)	23	上限値(二要因).txt	主効果による等
		平均値の区間推定(母分散既知)	4	手順1.doc	手順2.doc 等
		平均値の区間推定(母分散未知)	3	手順1.doc	手順3.doc 等
		比率の区間推定	4	手順1.doc	手順2.doc 等
		相関係数の検定	2	手順1.doc	手順2.doc 等
		1標本の平均値の検定	2	手順1.doc	手順2.doc 等
		1標本の比率の検定	2	手順1.doc	手順2.doc 等
		2標本の平均値の差の検定(対応のあるケース)	2	手順1.doc	手順2.doc 等
		2標本の平均値の差の検定(対応の無いケース)	2	手順1.doc	手順2.doc 等
	プログラム	1標本の平均値の検定.xls			
		2標本の平均値の検定(対応の無いケース).xls			
		2標本の検定(対応のあるケース).xls			
		カイ2乗検定.xls			
		一要因分散分析(被験者内).xls			
		一要因分散分析(被験者間).xls			
		二要因分散分析.xls			
		二要因分散分析(被験者内外).xls			
		区間推定.xls			
		標本比率の検定.xls			
		相関係数の検定.xls			
標本抽出実験	インストール クション	クロス表の関連度	11	手順1.doc	手順2.doc 等
		不偏分散の分布	4	手順1.doc	手順2.doc 等
		標本分散の分布	3	手順1.doc	手順2.doc 等
		標本平均の分布(t分布)	7	手順1.doc	手順2.doc 等
		標本平均の理論分布	4	手順1.doc	手順2.doc 等
		標本平均値の分布	3	手順1.doc	手順2.doc 等
		標本平均値の差の分布	3	手順1.doc	手順2.doc 等
		標本比率の分布	3	手順1.doc	手順2.doc 等
		標本比率の差の分布	3	手順1.doc	手順2.doc 等
		標本比率の理論分布	4	手順1.doc	手順2.doc 等
		正規分布	1	概要.doc	
		相関係数	4	概要.doc	概要2.doc 等
		3標本の平均値	18	手順1.doc	手順10.doc 等
		t分布	1	概要.doc	
	プログラム	クロス表の関連度.xls			
		不偏分散の分布.xls			
		標本分散の分布.xls			
		標本平均の分布(t分布).xls			
		標本平均値の分布.xls			
		標本平均値の差の分布.xls			
		標本平均値の理論分布.xls			
		標本比率の分布.xls			
		標本比率の差の分布.xls			
		標本比率の理論分布.xls			
		正規分布.xls			
		相関係数の分布.xls			
		3標本の平均値の分布.xls			
		t分布.xls			
無作為抽出 の考え方	インストール プログラム	標本平均	3	手順1.doc	手順2.doc 等
		標本比率	3	手順1.doc	手順2.doc 等
		標本抽出シミュレーション(平均値).xls			
		標本抽出シミュレーション(比率).xls			

第2章 各学習プログラムでの課題と実施手順

本章では、学習プログラムごとに、そこで学ぶべき内容・課題とその実施手順を示す。本章で紹介する学習プログラムは以下の通りである。

◆ 標本抽出実験

- (1) 母集団からの無作為標本の抽出
- (2) 標本平均値の分布
- (3) 標本平均値の理論分布
- (4) 標本比率の分布
- (5) 標本比率の理論分布
- (6) 標本分散の分布
- (7) 不偏分散の分布
- (8) 2標本の平均値の差の分布
- (9) 正規分布の作図
- (10) 不偏分散を用いた標準化標本平均の分布
- (11) t分布の作図
- (12) 3標本の平均値の分布
- (13) 標本相関係数の分布
- (14) クロス表の関連の強さの指標 (χ^2 統計量)

◆ 区間推定・統計的仮説検定

- (1) 母分散が既知の場合母平均の区間推定
- (2) 母分散が未知の場合母平均の区間推定
- (3) 母比率の区間推定
- (4) 1標本の平均値に対する検定 (母分散未知)
- (5) 2標本の平均値の差に対する仮説検定
(母分散は未知だが、等しいことを仮定できる場合)
- (6) 2標本の平均値の差に対する仮説検定 (2標本が対応している場合)
- (7) 1標本の母比率に対する仮説検定
- (8) 2標本の母比率の差に対する仮説検定
- (9) χ^2 適合度検定
- (10) 相関係数の検定
- (11) 一要因の分散分析
- (12) 二要因の分散分析
- (13) クロス表の関連度についての χ^2 検定

2. 1 標本抽出実験

(1) 母集団からの無作為標本の抽出

◆プログラム

フォルダ：標本抽出実験¥無作為抽出の考え方¥プログラム¥

ファイル名：

①標本抽出シミュレーション（平均値）.XLS

②標本抽出シミュレーション（比率）.XLS

◆目的

仮想的な母集団に対して無作為標本の抽出実験を行う。得られた標本データから標本平均値（標本比率）を求め、それがどのような値をとるのかを調べる。

◆検討課題

標本統計量（平均、比率）に関する以下の特性を確認する。

①標本平均（比率）は母平均（母比率）と一致するとは限らない

②標本平均（比率）は常に同じ値をとるわけではなく、標本の取り出し方によってその値は異なる。

◆実験手順

「A市における自宅でのインターネット利用状況」調査

調査の①	調査対象②	実験の目的
母集団④	者数 500	
分布を調べる	利用者 239	
	非利用者 261	
⑥	利用率 47.8	手順
		結果の削除

試行回数	1	③ 2				
利用者	250	239				
非利用者	250	261				
利用率	50.0	47.8				

① 調査の実施

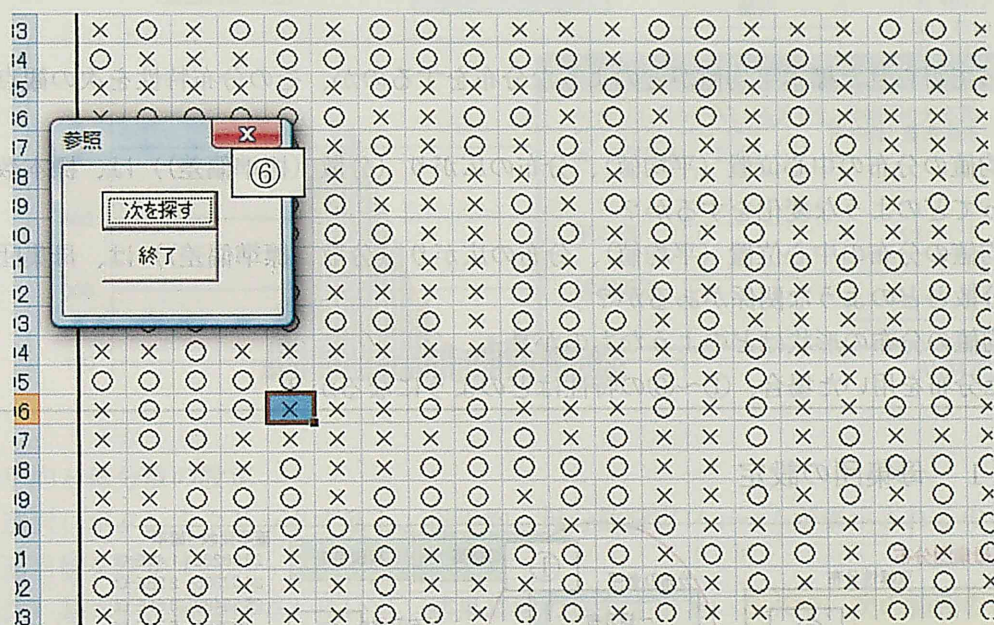
「調査の実施」ボタン（①）をクリックすると、②の欄に表示されている調査対象者数だけの無作為標本が母集団から抽出され、その内訳人数（インターネット利用者、非利用者）と利用率も同じく②の欄に表示される。さらに、「調査の実施」ボタンをクリックすると、先の回とは異なった標本が抽出され、その結果も②の欄に表示される。このような複数回の標本抽出の結果は、③の欄に付加されていくので、それらの値を比較すれば、標本比率がどのような値をとるかについて調べることができる。なお、調査対象者数を変化させてシミュレーション

ョンを行うことも可能で、その場合には、調査対象者数欄に直接入力すればよい。(S)

② 母集団の参照

具体的に母集団（60 万の仮想集団）からの標本抽出の状況を見るためには、「母集団参照」ボタン（④）をクリックすればよい。これにより以下のような「母集団」シートの内容が表示される。母集団シートの各セルは 1 人の市民に対応しており、その市民のインターネット利用状況は○（利用）、×（非利用）によって表示されている。従って、母比率が 0.5 であることから○セル、×セルともにその数は 30 万である。

また、青で着色されているセルは、標本抽出された者を表しており、従って着色セル数は調査対象者数となる。そして、着色セル中の○数、×数が抽出標本におけるインターネット利用者数、非利用者数となる。なお、「参照」フォームの「次を探す」ボタン（⑥）をクリックすれば、次の着色セルに移動することができる。



13	×	○	×	○	○	×	○	○	×	×	×	×	○	○	×	×	×	○	○	×
14	○	×	×	×	○	○	○	○	○	○	×	○	○	○	○	×	×	×	○	○
15	×	×	×	×	○	×	×	×	×	○	×	○	×	○	×	×	×	×	×	○
16	×	×	×	×	○	×	×	×	×	○	×	○	×	○	×	×	×	×	×	○
17	参照	×	×	×	×	○	×	×	×	○	×	○	×	×	×	×	×	×	×	○
18	次を探す	×	×	×	×	○	×	×	×	○	×	○	×	×	×	×	×	×	×	○
19	終了	○	○	×	×	○	×	×	×	○	×	○	×	○	×	×	×	×	×	○
20		○	○	×	×	○	×	×	×	○	×	○	×	○	×	×	×	×	×	○
21		○	○	×	×	○	×	×	×	○	×	○	×	○	×	×	×	×	×	○
22		×	×	×	×	○	×	×	×	○	×	○	×	×	×	×	×	×	×	○
23		○	○	○	×	×	×	×	×	○	×	○	×	○	×	×	×	×	×	○
24		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	○
25		○	○	○	×	×	×	×	×	○	×	○	×	○	×	×	×	×	×	○
26		×	○	○	×	×	×	×	×	○	×	○	×	○	×	×	×	×	×	○
27		×	○	○	×	×	×	×	×	○	×	○	×	○	×	×	×	×	×	○
28		×	×	×	×	○	×	×	×	○	×	○	×	×	×	×	×	×	×	○
29		×	×	○	○	×	○	○	○	×	×	×	×	×	×	×	×	×	×	○
30		○	○	○	○	○	○	○	○	×	×	×	×	×	×	×	×	×	×	○
31		×	×	×	○	×	○	×	○	○	×	○	×	○	×	○	×	×	×	○
32		○	○	○	×	×	×	×	×	○	×	○	×	○	×	×	×	×	×	○
33		×	○	○	×	×	×	○	○	×	○	×	○	×	×	×	×	×	×	○

③ 分布を調べる

標本比率がどのような分布を示すかについては、標本抽出実験を多数回繰り返し、そこで得られた標本比率の分布を調べれば、そのおおよその傾向を把握することができる。「分布を調べる」ボタン（⑤）そのためのもので、これにより標本比率の分布シートがアクティブとなり、多数の抽出実験を行うことができる。

(2) 標本平均値の分布

◆プログラム

フォルダ名：標本抽出実験×標本統計量の分布×プログラム×

ファイル名：標本平均値の分布.XLS

◆目的

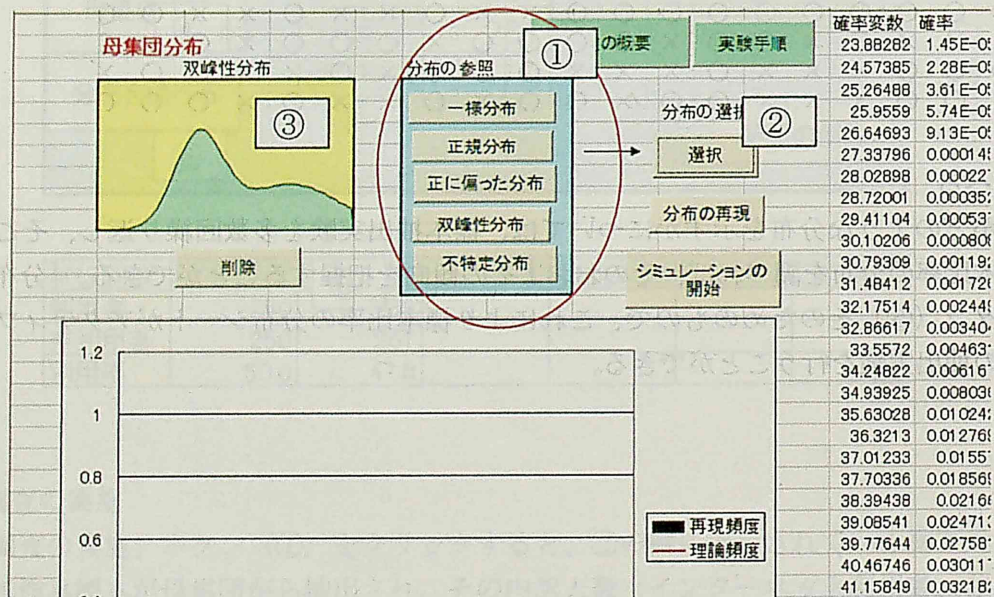
標本平均値は、母集団からの取り出し方によって異なった値をとる。ここでは、パソコン上での標本抽出実験を多数回行い、その結果得られた多数の標本平均値の分布を調べることににより、一般に標本平均がどのような分布をとるのかを検討する。母集団としては、一様分布、正規分布、正に偏った分布、双峰性分布、さらには一般的に形状を特定できない分布（不特定分布）を設定することが可能で、母集団の分布の違いによる標本平均値の分布の違いについても検討することができる。

◆検討課題

実験によって得られた標本平均値がどのような分布をするのか、その分布特性を次の観点から調べる。

- ①標本平均値の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、標本数の変化によってどのような変化をするか？
- ②標本平均値の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、母集団の平均値、分散とどのような関係があるか？
- ③標本平均値の分布の形状はどのようなものか？
- ④異なった分布を用いた場合、①～③の特性はどのようなになるか？

◆実験手順 1 母集団の設定



①分布の参照

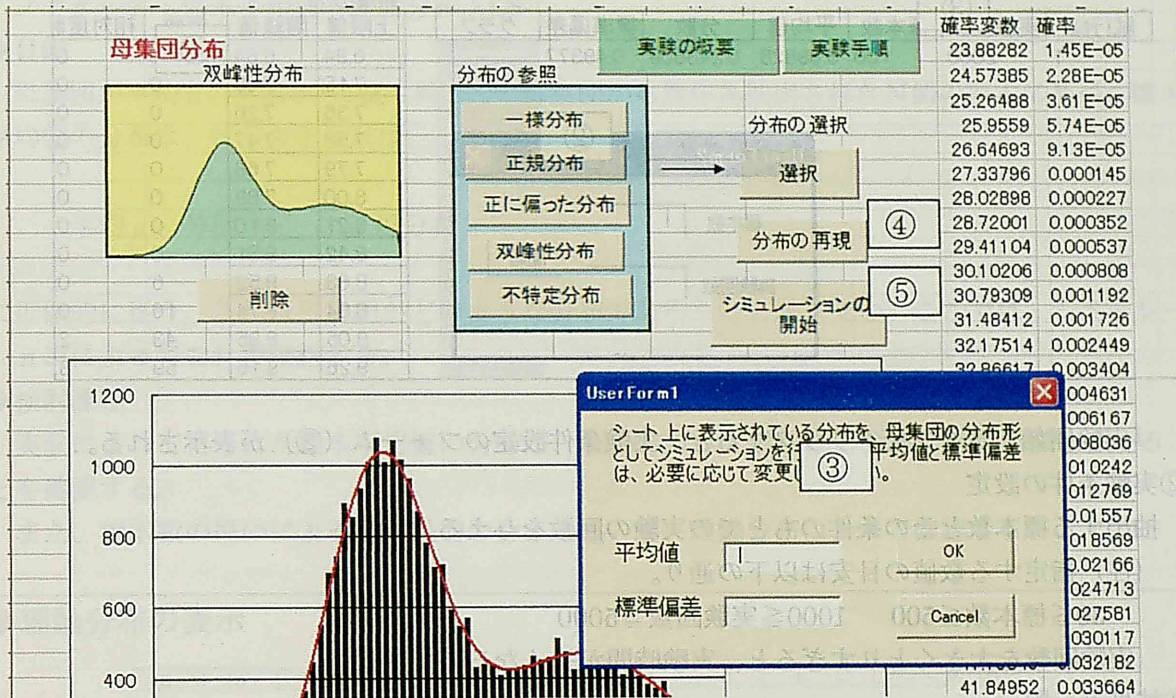
* 一様分布～不特定分布の中から1つを選択し、該当ボタンをクリックする

* 分布形がシート上に表示される (③)

*正に偏った分布、双峰性分布、不特定分布については、該当ボタンを繰り返しクリックすれば、形状の異なる分布が得られるので、その中から適当な分布を選択する

②分布の選択

左図に表示された分布でシミュレーションを行う場合には、「選択」ボタンをクリックする



③母平均と母分散の設定

「選択」ボタンをクリックすると、母平均、母分散設定フォームが表示されるので、該当欄に適当値を設定する。

(注) 基本的に値は何でも良いが、結果表示の見やすさという点で、次の範囲の値を設定するのが良い

$$10 \leq \text{母平均値} \leq 100 \quad 10 \leq \text{母分散} \leq 100$$

④分布の再現

乱数を用いたシミュレーションによって、想定した母集団分布が再現されるか確認する。ここでは、30000のデータを発生させ、それによって得られる分布を図として描く。

(注) 確認のみが目的なので、かならずしも実行する必要はない

⑤シミュレーション開始

「シミュレーション開始」ボタンをクリックすると、実験シートが追加される。実験シートには「分布形名一連番」の名称が与えられる。例えば、「双峰性分布」であれば、シート名は「双峰性-1」等となる。

◆手順2 無作為抽出実験

母集団分布: 双峰性分布							頻度分布										
平均	10	分散	25	標準偏差	5	試行No	実験回数	標本数	平均値	分散	標準偏差	グラフ	上限値	階級値	度数	相対度数	
							1	2000	100	9.99828	0.243809	0.49377		6.94	6.84	0	0
													7.15	7.05	0	0	
													7.36	7.26	0	0	
													7.58	7.47	0	0	
													7.79	7.68	0	0	
													8.00	7.89	0	0	
													8.21	8.10	0	0	
													8.42	8.31	0	0	
													8.63	8.52	6	0	
													8.84	8.74	16	0	
													9.05	8.95	43	2	
													9.26	9.16	59	3	

①
 実験開始

分析結果クリア

③

UserForm2
 ②
 標本数
 実験回数
 開始

「実験開始」ボタンをクリックすると、実験条件設定のフォーム(②)が表示される。

②実験条件の設定

抽出する標本数とその条件のもとでの実験の回数を与える。

(注) 指定する数値の目安は以下の通り。

$30 \leq \text{標本数} \leq 500$ $1000 \leq \text{実験回数} \leq 5000$

実験回数を大きくとりすぎると、実験時間が長くなるので注意。

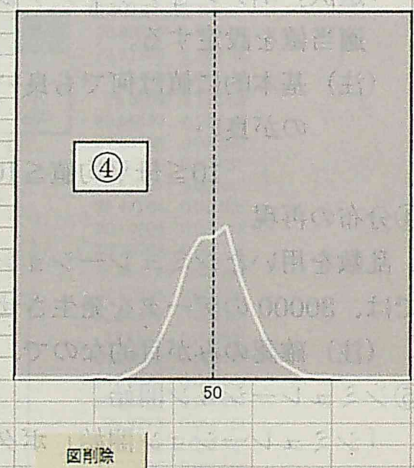
③結果表示欄

実験結果はこの欄に追加される。実験条件をシステマチックに変更して複数回実施する。

④頻度分布

度数分布表とそれをヒストグラムで表現した右のような図が作成される。実験回数が少ない場合には、分布形が明確に現れない場合があるので注意。分布形は細部にこだわらず大まかな印象で捉えると良い。

		頻度分布			
標準偏差	グラフ	上限値	階級値	度数	相対度数
1.00109		43.89	43.68	0	0.0
		44.31	44.10	0	0.0
		44.73	44.52	0	0.0
		45.15	44.94	0	0.0
		45.57	45.36	0	0.0
		45.99	45.78	0	0.0
		46.42	46.21	1	0.0
		46.84	46.63	2	0.1
		47.26	47.05	0	0.0
		47.68	47.47	16	0.5
		48.10	47.89	48	1.6
		48.52	48.31	130	4.3
		48.95	48.74	228	7.6
		49.37	49.16	381	12.7
		49.79	49.58	456	15.2
		50.21	50.00	461	15.4
		50.63	50.42	496	16.5
		51.05	50.84	337	11.2
		51.48	51.26	222	7.4
		51.90	51.69	125	4.2
		52.32	52.11	58	1.9
		52.74	52.53	30	1.0



図削除

(3) 標本平均値の理論分布

標本平均値の理論分布 (4)

◆プログラム

フォルダ名：標本抽出実験×標本統計量の分布×プログラム×

ファイル名：標本平均値の理論分布.XLS

◆目的

母集団が平均 μ 、分散 σ^2 に従う時、この母集団から無作為抽出されたN個の標本に基づく標本平均値の分布は、Nが大きければ、

平均 μ 、分散 $\frac{\sigma^2}{N}$ の正規分布

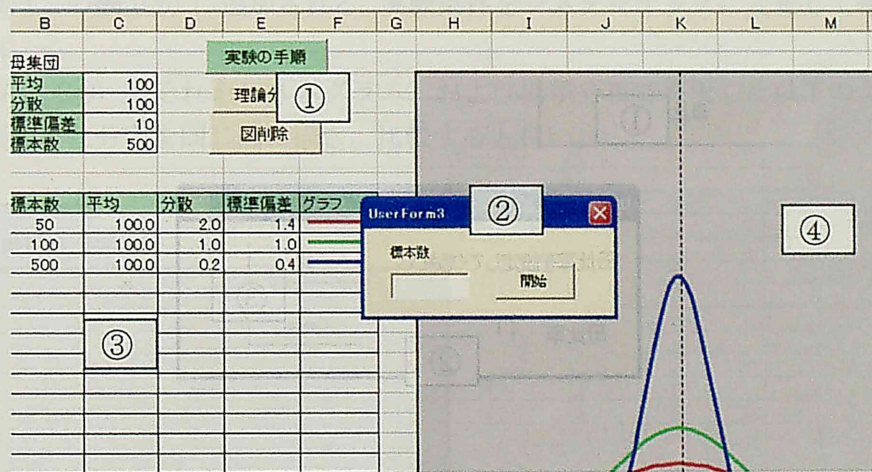
に近似的に従う。これは母集団がどのような分布であろうとも成り立つ。ここでは、シミュレーションによってこれを確認する。

◆検討課題

実験によって得られた標本平均値の分布が、理論によって示される分布と良く類似していることを確認する。

また、標本数の違いによる標本平均の分布の違いを、理論分布によって確認する。

◆理論分布の表示



①実験の開始

「実験開始」ボタンをクリックすると、実験条件設定のフォーム(②)が表示される。

②実験条件の設定

抽出する標本数を設定する。

(注) 指定する数値の目安：30 ≤ 標本数 ≤ 1000

③結果表示欄

実験結果はこの欄に追加される。標本数を変えて複数の実験を行い、標本数の違いが分布の形状(散布度)に与える影響を確認する。

④頻度分布：標本平均の理論分布がヒストグラムで表現される(④)。

(4) 標本比率の分布

市役所長の選挙の模擬実験 (8)

◆プログラム

フォルダ名：標本抽出実験¥標本統計量の分布¥プログラム¥

ファイル名：標本比率の分布.XLS

◆目的

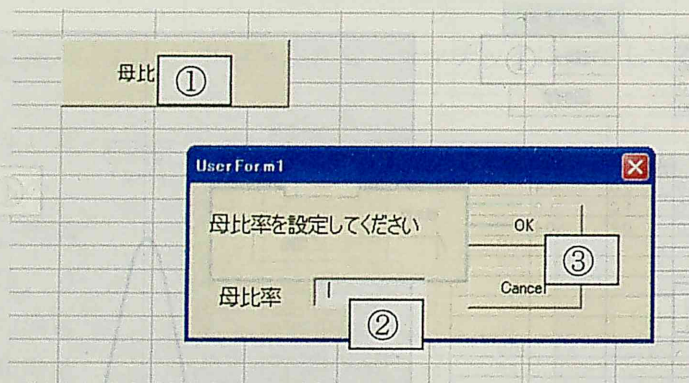
標本比率は、母集団からの取り出し方によって異なった値をとる。ここでは、パソコン上での標本抽出実験を多数回行い、その結果得られた多数の標本比率の分布を調べることにより、一般に標本比率がどのような分布をとるのかを検討する。

◆検討課題

実験によって得られた標本比率平均値がどのような分布をするのか、その分布特性を次の観点から調べる。

- ①標本比率の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、標本数の変化によってどのような変化をするか？
- ②標本比率の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、母比率とどのような関係があるか？
- ③標本比率の分布の形状はどのようなものか？

◆実験手順



①母比率の設定

「母比率の設定」ボタンをクリック（①）すると、母比率設定フォームが表示されるので、該当欄（②）に適当値を設定する。

（注）母比率は0～1の範囲で設定する

②シミュレーション開始 母比率

設定画面で「OK」ボタン（③）をクリックすると、以下の実験シートが追加される。実験シートには「標本比率-連番」の名称が与えられる。

母集団分布							頻度分布					国府野
母比率	0.5	実験開始		分析結果クリア			上限値	階級値	度数	相対度数	累積	相対度数
母分散	0.25						0.005	0.00				
母標準偏差	0.5						0.03	0.02				
試行No	実験回数	標本数	平均値	分散	標準偏差	グラフ	0.05	0.04				
							0.07	0.06				
							0.09	0.08				
							0.11	0.10				
							0.13	0.12				
							0.15	0.14				
							0.17	0.16				
							0.19	0.18				
							0.21	0.20				
							0.23	0.22				
							0.25	0.24				

母集団分布

母比率	0.5
母分散	0.25
母標準偏	0.5

①

実験開始

分析結果クリア

試行No	実験回数	標本数	平均	Us	②	2
③						

標本数

実験回数

開始

頻度

上限

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

0.0

①実験の開始

「実験開始」ボタン (①) をクリックすると、実験条件設定のフォーム (②) が表示される。

②実験条件の設定

抽出する標本数とその条件のもとでの実験の回数を与える。

(注) 指定する数値の目安は以下の通り。

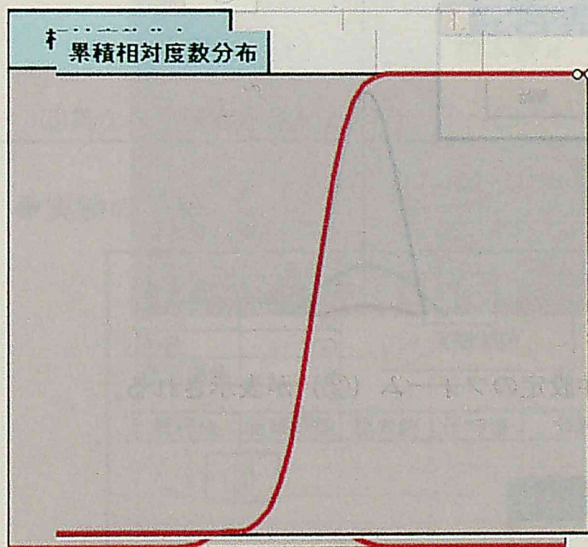
$50 \leq \text{標本数} \leq 500$ $3000 \leq \text{実験回数} \leq 10000$

③結果表示シート

実験結果はこの欄 (③) に追加される。実験条件をシステマチックに変更して複数回実施する。

④頻度分布

度数分布表 (④) とそれをヒストグラム、および累積分布で表現した以下のような図が作成される。異なった分布を同じ図に描き、比較すると良い。



(5) 標本比率の理論分布

◆プログラム

フォルダ名：標本抽出実験×標本統計量の分布×プログラム×

ファイル名：標本比率の理論分布.XLS

◆目的

母比率が π である母集団から N 個の標本に基づく標本比率の分布は、 N がある程度大きければ、

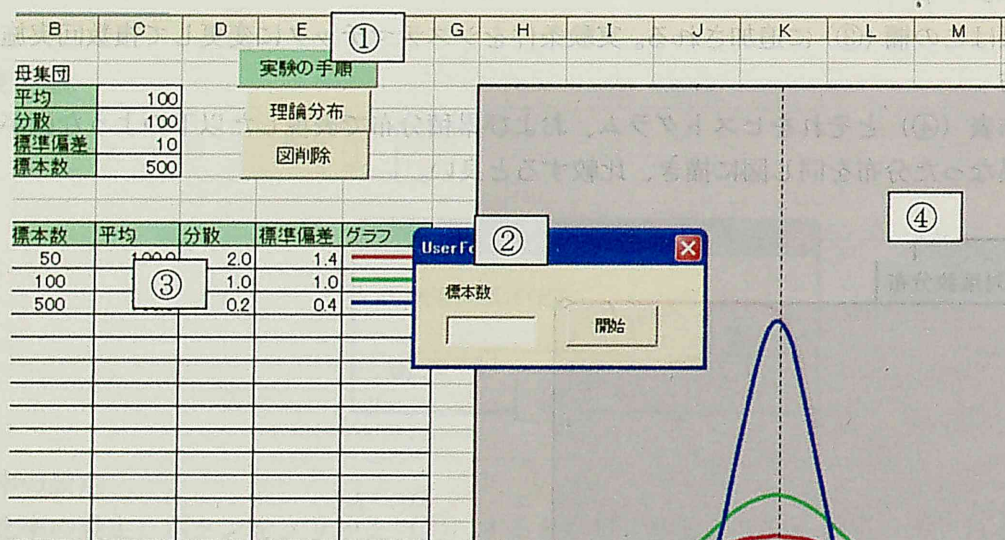
平均 π 、分散 $\frac{\pi \times (1 - \pi)}{N}$ の正規分布

で近似できる。ここでは、理論分布がシミュレーションによる標本比率の分布をかなり正確に近似し得ていることを確認する。

◆検討課題

実験によって得られた標本比率の分布が、理論によって示される分布と良く類似していることを確認する。また、標本数によって標本比率の分布（散布度）がどのように変化するかを、理論分布によって調べる。

◆理論分布の表示



①実験の開始

「実験開始」ボタンをクリックすると、実験条件設定のフォーム（②）が表示される。

②実験条件の設定

抽出する標本数を設定する。

（注）指定する数値の目安は以下の通り。

$30 \leq \text{標本数} \leq 1000$

③結果表示欄

実験結果はこの欄に追加される。標本数を変えて複数の実験行い、標本数の違いが分布の形状（散布度）に与える影響を確認する。

④頻度分布：標本比率の理論分布がヒストグラムで表現される（④）。

(6) 標本分散の分布

◆プログラム

フォルダ名：標本抽出実験×標本統計量の分布×プログラム×

ファイル名：標本分散の分布.XLS

◆目的

標本分散は、標本平均等と同様、母集団からの標本の取り出し方によって異なった値をとる。ここでは、パソコン上での標本抽出実験を多数回行い、その結果得られた多数の標本分散の分布を調べることににより、一般に標本分散がどのような分布をとるのかを検討する。

◆検討課題

標本分散の分布の特性を次の観点から調べる

- ①標本分散の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、標本数の変化によってどのような変化をするか？
- ②標本分散の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、母集団の分散とどのような関係があるか？
- ③標本分散の分布の形状はどのようなものか？
- ④標本平均の周りのバラツキ（標本分散）ではなく、母平均の周りのバラツキ（母平均を用いた分散）はどのような値をとるか？

$$\text{標本分散} = \frac{\sum_i (x_i - \bar{x})^2}{n}, \quad \bar{x} = \frac{\sum_i x_i}{n}$$

$$\text{母平均を用いた分散} = \frac{\sum_i (x_i - \mu)^2}{n}$$

但し 観測値 x_1, x_2, \dots, x_n 標本数 n 標本平均 \bar{x} 母平均 μ

- ⑤異なった分布を用いた場合、①～③の特性はどのようなようになるか？

◆実験の手順

母集団分布: 正規分布		①		⑤	
平均	50	実験開始		分析結果クリア	
分散	100			母平均を用いた分散	
標準偏差	10				
試行No	実験回数	標本数	平均値	分散	標準偏差
③					
		UserForm2		頻度分布	
		②		上限値	階級値
		標本数		3.33	0.00
		実験回数		10.00	6.67
		開始		16.67	1
				23.33	2
				30.00	26.67
				36.67	33.33
				43.33	40.00
				50.00	46.67
				56.67	53.33
				63.33	60.00

①実験の開始

「実験開始」ボタンをクリックすると、実験条件設定のフォーム (②) が表示される。

②実験条件の設定

抽出する標本数とその条件のもとでの実験の回数を与える。

(注) 指定する数値の目安は以下の通り。

$$10 \leq \text{標本数} \leq 500 \quad 1000 \leq \text{実験回数} \leq 5000$$

標本数はこれまでのシミュレーション異なり、小さなケースも実行すること

(注) 実験回数を大きくとりすぎると、実験時間が長くなるので注意。

③結果表示シート

実験結果はこの欄に追加される。実験条件をシステムチックに変更して複数回実施する。

④頻度分布

度数分布表とそれをヒストグラムで表現した図が作成される

⑤母平均を用いた場合

標本分散は以下のように標本平均の周りのバラツキを表す統計量である。これに対して、母平均の周りのバラツキはどのようにになるのか、そしてその分布は標本分散の分布とどのように異なるかを調べる。

$$\text{標本分散} = \frac{\sum_i (x_i - \bar{x})^2}{n} \quad , \quad \text{母平均を用いた分散} = \frac{\sum_i (x_i - \mu)^2}{n}$$

◆プログラム

フォルダ名：標本抽出実験¥標本統計量の分布¥プログラム¥

ファイル名：不偏分散の分布.XLS

◆目的

不偏分散、標本分散、そして母平均を用いた分散はそれぞれどのような分布となるのか。ここでは、パソコン上での標本抽出実験を多数行い、その結果得られた各種分散の分布を比較することにより、その性質の違いを検討する。

◆ 検討課題

不偏分散の分布の特性を次の観点から調べる

- ①不偏標本分散の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、標本数の変化によってどのような変化をするか？
- ②不偏分散の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、母集団の分散とどのような関係があるか？
- ③標本分散の分布の形状はどのようなものか？ それは不偏分散とどのように異なるのか？
- ④標本分散ではなく、母平均を用いた分散はどのような値をとるか？ それは不偏分散とどのように異なるのか？
- ⑤異なった分布を用いた場合、①～④の特性はどのようにになるか？

◆実験の手順

	B	C	D	E	F	G	H	I	J	K	L	M
集団分布: 正規分布				①	実験の目的	実験の手順						
平均値		50										
分散		100		不偏分散		母平均	⑤		標本分散		分析結果クリア	⑦
標準偏差		10										

行No	実験回数	標本数	平均値	分散	分散種別	グラフ
	③					

頻度分布		上限値	階級値	度数	相対度数
		3.33	0.00		
		10.00	6.67		
		16.67	13.33		
		23.33	20.00		
		30.00	26.67		
		36.67	33.33		
		43.33	40.00		
		50.00	46.67		
		56.67	53.33		
		63.33	60.00		

①不偏分散

「不偏分散」ボタンをクリックすると、実験条件設定のフォーム (②) が表示される。

②実験条件の設定

抽出する標本数とその条件のもとでの実験の回数を与える。

(注) 指定する数値の目安は以下の通り。

10 ≤ 標本数 ≤ 500 1000 ≤ 実験回数 ≤ 5000

標本数はこれまでのシミュレーション異なり、小さなケースも実行すること

(注) 実験回数を大きくとりすぎると、実験時間が長くなるので注意。

③結果表示シート

実験結果はこの欄に追加される。実験条件をシステムチェックに変更して複数回実施する。

④頻度分布

度数分布表とそれをヒストグラムで表現した図が作成される

⑤母平均を用いた場合

標本分散は以下のように標本平均の周りのバラツキを表す統計量である。これに対して、母平均の周りのバラツキはどのようになるのか、そしてその分布は標本分散の分布とどのように異なるかを調べる。

$$\text{不偏分散} = \frac{\sum_i (x_i - \bar{x})^2}{n-1}, \quad \text{母平均を用いた分散} = \frac{\sum_i (x_i - \mu)^2}{n}$$

⑥標本分散

標本分散は、平均からの偏差の 2 乗を標本数で割ったものである。この標本分散と不偏分散の分布がどのように異なるかを調べる。

(8) 2標本の平均値の差の分布

◆プログラム

フォルダ名：標本抽出実験¥標本統計量の分布¥プログラム¥

ファイル名：標本平均値の差の分布.XLS

◆目的

2つの母集団からそれぞれ無作為に抽出した2つの標本からは、2つの標本平均値を得ることができるが、その差は、母集団からの取り出し方によって異なった値をとる。ここでは、パソコン上での標本抽出実験を多数回行い、その結果得られた多数の標本平均値の差の分布を調べることにより、一般に標本平均の差がどのような分布をとるのかを検討する。

◆検討課題

標本平均値の差の分布の特性を次の観点から調べる

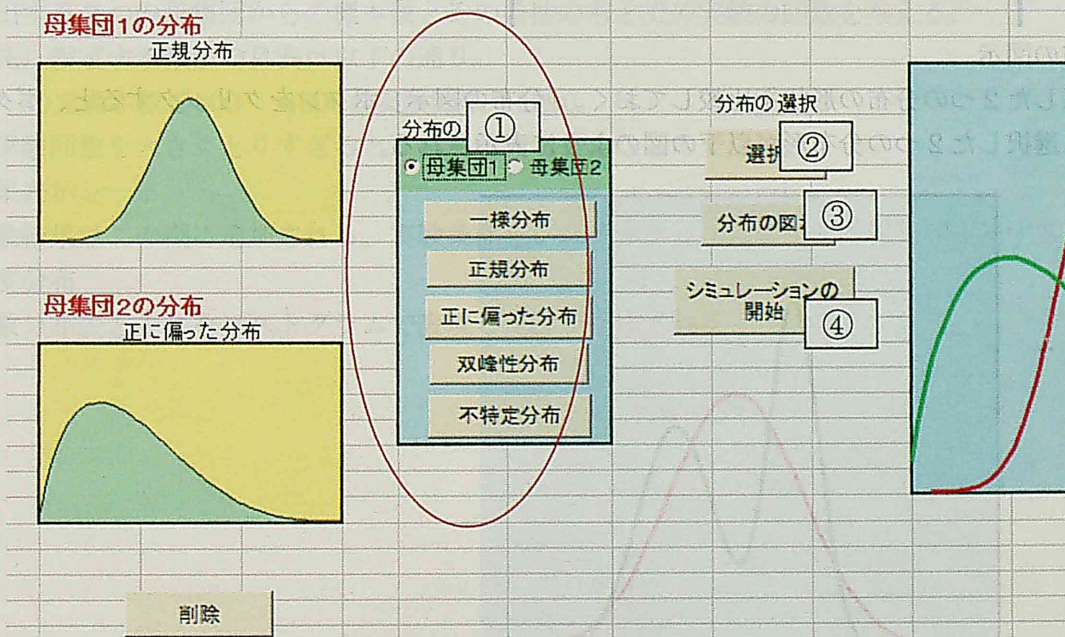
- ①標本平均値の差の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、標本数の変化によってどのような変化をするか？
- ②標本平均値の差の分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、母集団の平均値、分散とどのような関係があるか？
- ③標本平均値の差の分布の形状はどのようなものか？
- ④異なった分布を用いた場合、①～③の特性はどのようなになるかを、以下の点から検討する。

◆2つの母集団の分布が同じ場合、異なった場合

- ・分布形が同じ場合、異なった場合
- ・2つの母集団の母平均が等しい場合、異なった場合
- ・2つの母集団の母分散が等しい場合、異なった場合

◆2つの母集団からの標本数が同じ場合、異なった場合 等

◆実験手順その1



①分布の参照

* 分布形を指定する母集団を、OptionButton を用いて選択する

(母集団1 Or 母集団2)

* 一様分布～不特定分布の中から1つを選択し、該当ボタンをクリックする

* 分布形がシート上に表示される

* 正に偏った分布、双峰性分布、不特定分布については、該当ボタンを繰り返しクリックすれば、形状の異なる分布が得られるので、その中から適当な分布を選択すればよい

* 以上の手順を2つの母集団について行う。

②分布の選択と母平均、母分散の設定

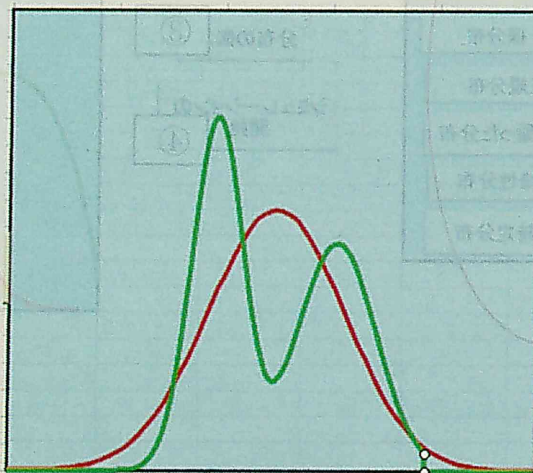
選択した分布でシミュレーションを行う場合には、「選択」ボタンをクリックする。「選択」ボタンをクリックすると、以下の母平均、母分散設定フォームが表示されるので、該当欄に適当値を設定する。

(注) 基本的に値は何でも良いが、結果表示の見やすさという点で、次の範囲の値を設定するのが良い

$10 \leq \text{母平均値} \leq 100$ $10 \leq \text{母分散} \leq 100$

③分布の図示

選択した2つの分布の形状を比較しておく。「分布の図示」ボタンをクリックすると、ボタンの右側に選択した2つの分布形が以下の図のように表示される。



④シミュレーション開始ボタンをクリックする

◆実験の手順その2

図表の赤矢印 (9)

①シミュレーション開始

「分布の選択」シートで、「シミュレーション開始」ボタンをクリックすると、以下の実験シートが追加される。実験シートには「平均値の差-連番」の名称が与えられる。

B	C	D	E	F	G	H	I	J	K	L	M	N	O
母集団1		母集団2			②								
分布	正規分布	分布	正に偏った分布										
平均	50	平均	60										
分散	100	分散	100		実験開始	分析結果クリア							
標準偏差	10	標準偏差	10										

試行No	④	数	標本数	平均値	分散	標準偏差	グラフ	頻度分布	⑤	数
								上限値	階級値	度数
								-22.23	-22.65	
								-21.38	-21.81	
								-20.54	-20.96	
								-19.70	-20.12	
								-18.85	-19.28	
								-18.01	-18.43	
								-17.17	-17.59	
								-16.32	-16.75	
								-15.48	-15.90	
								-14.64	-15.06	
								-13.79	-14.22	
								-12.95	-13.37	
								-12.11	-12.53	
								-11.26	-11.69	
								-10.42	-10.84	
								-9.58	-10.00	
								-8.74	-9.16	
								-7.89	-8.31	
								-7.05	-7.47	

UserForm2

標本数

母集団1

母集団2

実験回数

開始

②実験の開始

「実験開始」ボタンをクリックすると、実験条件設定のフォーム (③) が表示される。

③実験条件の設定

抽出する2つの母集団からの標本数とその条件のもとでの実験の回数を与える。

(注) 指定する数値の目安は以下の通り。

$30 \leq \text{標本数} \leq 500$ $1000 \leq \text{実験回数} \leq 5000$

実験回数を大きくとりすぎると、実験時間が長くなるので注意。

④結果表示シート

実験結果はこの欄に追加される。実験条件をシステムチックに変更して複数回実施する。

⑤頻度分布

度数分布表とそれをヒストグラムで表現した図が作成される

(9) 正規分布の作図

◆プログラム

フォルダ名：標本抽出実験¥標本統計量の分布¥プログラム¥

ファイル名：正規分布.XLS

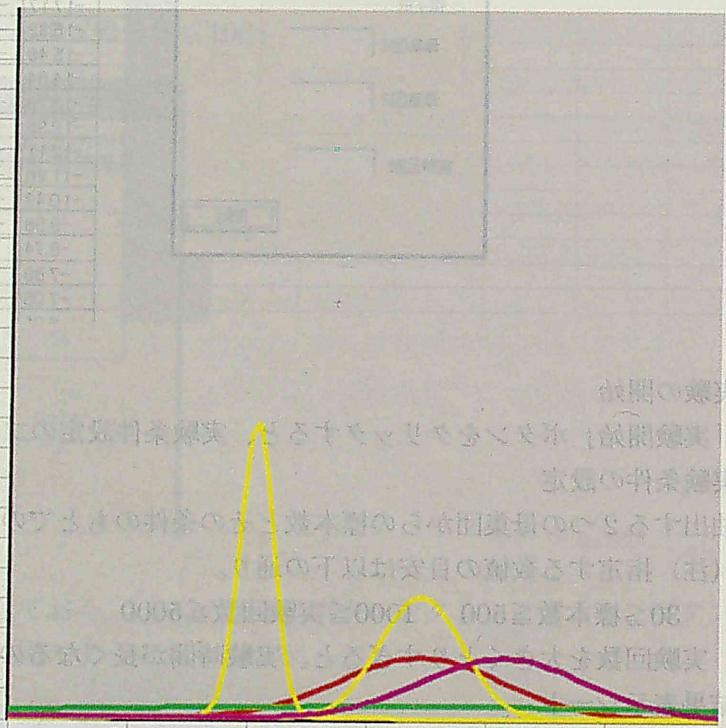
◆目的

正規分布がどのような形状を示すかを、その平均値、標準偏差を変化させながら調べる。但し、作図範囲の関係で、平均値、標準偏差は以下の範囲に限定している。

$$0 \leq \text{平均值} \leq 100$$

$5 \leq \text{標準偏差} \leq 50$

◆画面例

[illegible]

(10) 不偏分散を用いた標準化標本平均の分布

◆プログラム

フォルダ名：標本抽出実験¥標本統計量の分布¥プログラム¥

ファイル名：標本平均の分布(t分布).XLS

◆目的

標本平均を母分散で標準化した値（母分散標準化平均）と、不偏分散で標本平均値を標準化した値（不偏分散標準化平均）の分布特性の違いを次の観点から調べる。

- ①分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、標本数の変化によってどのような変化をするか？
- ②分布の中心位置（平均値）、分布の広がり（分散（標準偏差））は、母集団の平均値、分散とどのような関係があるか？
- ③分布の形状はどのようなものか？
- ④異なった分布を用いた場合、①～③の特性はどのようなようになるか？

◆実験の手順 その1

実験の目的			実験の手順			⑤		
分	50	00	10	①	母分散で標準化	種別	実験	理論
						分布	-	正規分布
						平均値		
						分散		
値	度数	相対度数	UserForm2			④		
	00		②					
③			標本数 <input type="text" value="1"/> 実験回数 <input type="text"/> <input type="button" value="開始"/>					
	40							
	20							
	00							
	80							
	60							
	40							
	20							
	00							
	80							

①母分散で標準化

「母分散で標準化」ボタンをクリックすると、実験条件設定のフォーム（②）が表示される。

②実験条件の設定

抽出する標本数とその条件のもとでの実験の回数を与える。

（注）指定する数値の目安は以下の通り。

$$5 \leq \text{標本数} \leq 100 \quad 2000 \leq \text{実験回数} \leq 5000$$

実験結果が滑らかな分布となるためには、実験回数を大きくした方がよい。

③結果表示シート

実験結果の度数分布表が③の欄に示され、さらにそれをヒストグラムで表現した図が④に作成される

⑤理論分布

理論分布として「標準正規分布」の度数分布とヒストグラムが、同じく、③と④に表示される。実験結果と理論分布が良く対応していることを確認する。

◆実験の手順 その2

The screenshot shows a software interface for statistical experiments. At the top, there's a menu bar with options like '実験の目的' (Purpose of Experiment), '実験の手順' (Experiment Procedure), '種別' (Type), '実験' (Experiment), and '理論' (Theory). Below this is a table with columns for '度数' (Frequency), '相対度数' (Relative Frequency), '標準正規' (Standard Normal), and 't分布' (t-distribution). A dialog box titled 'UserForm3' is open, showing input fields for '標本数' (Sample Size) and '実験回数' (Number of Experiments), and a '開始' (Start) button. Numbered callouts (1-7) indicate the sequence of operations: 1 points to the '不偏分散で標準化' (Standardize with unbiased variance) button; 2 points to the 'UserForm3' dialog box; 3 points to the '度数' column in the table; 4 points to the histogram area; 5 points to the '標準正規分布' (Standard Normal Distribution) button; 6 points to the 't分布' button; and 7 points to the '理論分布図削除' (Delete theoretical distribution graph) button.

①不偏分散で標準化

「不偏分散で標準化」ボタンをクリックすると、実験条件設定のフォーム(②)が表示される。

②実験条件の設定

抽出する標本数とその条件のもとでの実験の回数を与える。

(注) 指定する数値の目安は以下の通り。

$$5 \leq \text{標本数} \leq 100 \quad 2000 \leq \text{実験回数} \leq 5000$$

実験結果が滑らかな分布となるためには、実験回数を大きくした方が良い。

③結果表示シート

実験結果の度数分布表が③の欄に示され、さらにそれをヒストグラムで表現した図が④に作成される

⑤理論分布(標準正規分布)

理論分布として「標準正規分布」の度数分布とヒストグラムが、同じく、③と④に表示される。実験結果と理論分布との違いを詳細に検討する。

⑥理論分布(t分布)

理論分布として「t分布」の度数分布とヒストグラムが、同じく、③と④に表示される。実験結果は標準正規分布よりもt分布に類似した分布形を示すことを確認する。

⑦理論分布図の削除

標準正規分布、t分布を削除したい場合に用いる。

(11) t 分布の作図

赤いの箱は平均の本数 (S1)

◆プログラム

フォルダ名：標本抽出実験と標本統計量の分布とプログラム

ファイル名：t 分布.XLS

◆目的

自由度、つまり標本数を変えたとき、標本平均を不偏分散で標準化した値、 t がどのような分布を示すかを調べる。また、 t 分布と標準正規分布との違いについても調べる。

◆実験手順

The screenshot shows an Excel spreadsheet with a table for the experiment. The table has columns for '分布' (Distribution), '自由度' (Degrees of Freedom), and 'グラフ' (Graph). A dialog box titled 'User Form1' is overlaid on the spreadsheet, showing a text box for '自由度' (Degrees of Freedom) and a '開始' (Start) button. The spreadsheet also has a green box labeled '実験の目的と手順' (Purpose and Procedure of the Experiment) and a yellow box labeled '削除' (Delete).

分布	自由度	グラフ

- ① t 分布ボタンをクリックすると、自由度入力フォーム (②) が表示される。
- ② 自由度を入力し、開始ボタンをクリックすると、④の欄に分布形が表示される。自由度を変えた複数の試行によって、自由度の違いによる分布形の違いを検討することができる。
- ⑤ 標準正規分布
- ④の欄に標準正規分布の分布形が示され、 t 分布との違いを調べることができる。

(12) 3標本の平均値の分布

◆プログラム

フォルダ名：標本抽出実験×標本統計量の分布×プログラム×

ファイル名：3標本の平均値の分布.XLS

◆目的

全く同一の3つの母集団から抽出された3つの標本平均値間のバラツキを指標化（F値）し、その分布を調べる。さらに以下のような条件もとで、分布がどのように変わるかも調べる。

①標本数を変える

③母集団の数を、4以上とする

◆実験手順

B	C	D	E	F	G	H	I	J	K
母集団	第1	第2	第3	①		②		⑥	
分布	一様分布	一様分布	一様分布	実験準備		実験開始		理論分布	結
平均	50	50	50						
分散	100	100	100						
標準偏差	10	10	10						
実験の目的と手順									
相対度数分布									
階級設定									
階級範囲	階級値	10							
0~0.2	0.1	17.1							
0.2~0.4	0.3	15.6							
0.4~0.6	④ 0.5	10.6							
0.6~0.8	0.7	9.8							
0.8~1	0.9	8.8							
1~1.2	1.1	6.1							
1.2~1.4	1.3	4.7							
1.4~1.6	1.5	4.5							
1.6~1.8	1.7	3.9							
1.8~2	1.9	3.0							

UserForm4

③

標本数

3母集団で全て同一に設定

実験回数

開始

- ①実験準備ボタン（①）をクリックし、母集団の条件を設定する。
- ②実験開始ボタン（②）をクリックすると、③のフォームが表示されるので、標本数と実験回数を設定する。標本数、及び実験回数は以下の範囲で指定すると良い。
 $2 \leq \text{標本数} \leq 100$
 $1000 \leq \text{実験回数} \leq 5000$
- ③実験結果の度数分布表は④の欄に、図は⑤に表示される。
- ④理論分布ボタン（⑥）をクリックすると、その直前に行った実験と同様な場面におけるF値の理論分が、④、⑤の欄に表示される。

(13) 標本相関係数の分布

◆プログラム

フォルダ名：標本抽出実験¥標本統計量の分布¥プログラム¥

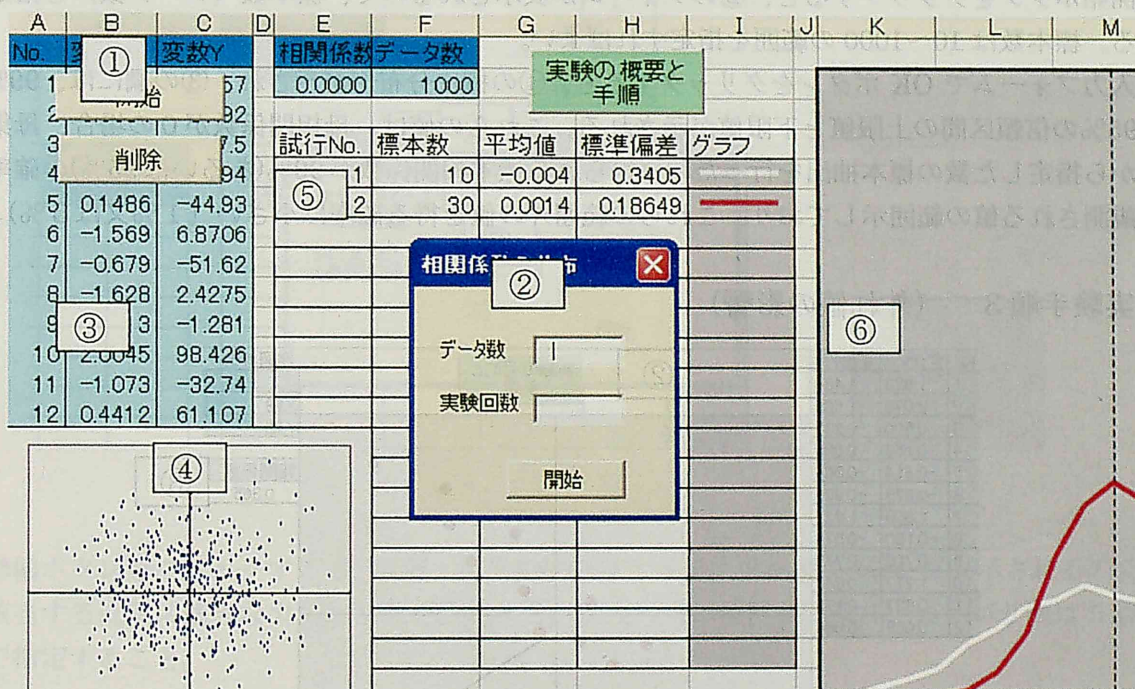
ファイル名：相関係数の分布.XLS

◆目的

相関係数の特性について、次の観点から学ぶ。

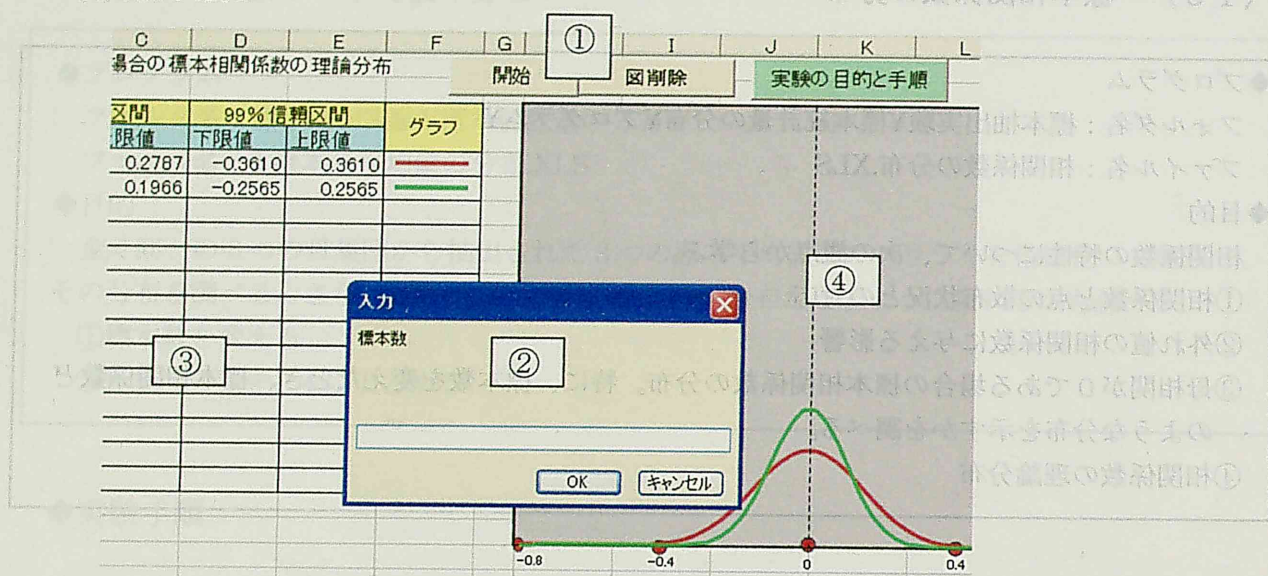
- ①相関係数と点の散布状況との対応
- ②外れ値の相関係数に与える影響
- ③母相関が0である場合の標本相関係数の分布。特に、標本数を変えたとき、標本相関係数どのような分布を示すかを調べる。
- ④相関係数の理論分布

◆実験手順 1



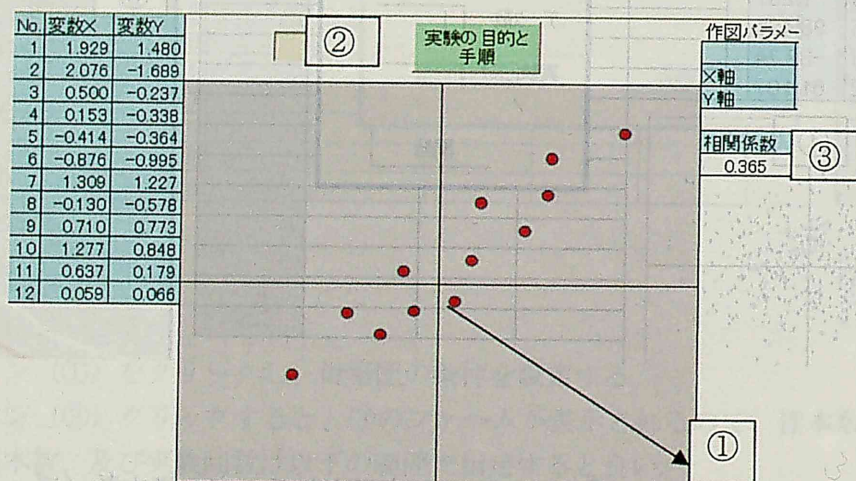
- ①開始ボタンをクリックすると、②のフォームが表示されるので、標本数（データ数）と実験回数を指定する。母集団はデータ数 1000（欄③）であり、④の散布図に示したように相関（母相関係数）は0である。標本数は10～200程度の範囲を指定する。実験回数は、滑らかな分布とするためには、少なくとも2000回以上を指定するのが良い。
- ②入力フォームで開始ボタンをクリックすると実験が開始され、得られた標本平均の分布の要約値が⑤の欄に、そして分布形が⑥の欄に表示される。標本数を変えた複数のシミュレーションを行い、標本数を増加させた場合の分布の変化を調べる。

◆実験手順 2



- ①開始ボタンをクリックすると、②のフォームが表示されるので、標本数（データ数）を指定する。標本数は10～1000の範囲を指定すればよい。
- ②入力フォームでOKボタンをクリックすると、④の欄に分布形が示され、③の欄には、99%、95%の信頼区間の上限値と下限値が示される。これらの値は、母相関係数が0の場合、母集団から指定した数の標本抽出を行った時に得られる標本相関係数が99%（あるいは95%）の確率で観測される値の範囲を示しており、これらの範囲外の値を得る確率は小さい（1%又は5%）。

◆実験手順 3 （外れ値の影響）



- ①12個の点のいずれか（複数でも可）を、全体の散布位置から離れた位置に移動させる。移動は、点の位置までマウス・ポインターをもっていき、左ボタンをクリックし、ボタンを押したまま目標の位置まで移動させればよい。
- ②修正ボタンをクリックすると、変化させた散布図に対応した相関係数値が③の欄に表示される。

(14) クロス表の関連の強さ指標 (χ^2 統計量)

◆プログラム

フォルダ名：標本抽出実験×標本統計量の分布×プログラム×

ファイル名：クロス表の関連度.XLS

◆目的

クロス表の関連度に関する基本的な考え方から、それを指標化した χ^2 統計量の意味を具体例から理解する。さらに、クロス表のカテゴリ数を変えることによって χ^2 統計量がどのように変わるかも調べる。

<実験手順> その1 カテゴリ数の設定

●標本抽出：クロス表の周辺度数比率と標本数を指定し、その母集団からの無作為標本に得られたクロス表から、カイ2乗統計量を求め、その分布を調べます

実験の目的と手順		標本数	実験回数	準備	開始
標本番号	カイ2乗値	試行回数			

②

UserForm6

第1変数のカテゴリ数 (表側変数) ③

第2変数のカテゴリ数 (表頭変数)

開始

- ①準備ボタンをクリックすると、表側、表頭変数のカテゴリ数指定画面(②)が表示されるので、該当する位置に適切なカテゴリ数を設定する。図表の表示位置の関係で、カテゴリ数は5以下で指定すること。

<実験手順> その2 カテゴリ比率の設定
 カテゴリ数指定画面で開始ボタンをクリックすると、④のカテゴリ比率入力フォームが表示されるので、該当欄に適当な値を設定する。変数ごとに比率の合計が1となるように注意すること。

<実験手順> その3 実験の開始

実験の目的 と手順		①		準備		<2変数が無関連な場合>			
		標本数	100	実験回数	3000	第1変数		第2変数	
				開始		cat-1		cat-2	
				②		cat-1	0.06	cat-2	0.09
						cat-2	0.08	cat-2	0.12
						cat-3	0.06	cat-2	0.09
						計	0.2	cat-2	0.3

標本 番号	カイ2 乗値	試行 回数	標本数	カテゴリ数		実験回 数	グラフ 表示
				第1変 数	第2変数		
1	9.226	1	100	3	4	3000	
2	4.612	2	100	④	4	3000	
3	9.924						
③ 4	3.296						
5	4.924						
6	6.965						
7	8.758						
8	3.721						
9	4.375						
10	10.443						
11	3.752						
12	2.671						
13	4.284						

①の欄の標本数と実験回数を指定した後、開始ボタン(②)をクリックする。これで実験が開始され、 χ^2 値のリストが③の欄に表示され、相対頻度分布が⑤に作図される。また、④の欄には実験条件が示される。

2 変数のカテゴリ数、カテゴリ比率を変化させ、それによる分布の違いを⑤の図によって検討する。

2. 2 区間推定・統計的仮説検定のためのツール

(1) 母分散が既知の場合の母平均の区間推定

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : 区間推定.XLS

シート名 : 母分散が既知

◆目的

得られた標本データに基づいて、母分散が既知の場合の母平均の信頼区間の推定を行う。

◆課題

このツールを用いて例題の解を求め、区間推定のプロセスを確認する。

◆手順

標本平均に基づいて母平均の区間推定を行う・・・母分散が既知の場合

クリア

データ入力 ①

標本平均	標本数
50	100

①区間推定を行うデータを設定する
(標本平均、母分散、標本数、信頼水準)

標本平均の分布 ②

分布形	平均
正規分布	μ

②標本平均の分布を特定し、分布形と分散を所定欄に設定する

分布を描く

95%の確率で標本平均が観測される区間 ③

③指定した信頼水準において標本平均が観測される区間を図示する

④標本平均が信頼区間の下限値に位置している場合を作図する

標本平均が95%下限値にある ④

分布を特定するために、平均値を μ だけ下にずらす ⑤

変換された標本平均の分布は？

分布形	平均	分散
正規分布	0	1

⑤分布を特定するために分布を μ (母平均) だけ負の方向にずらし、その時の標本平均の分布を特定する

母平均の最大値の推定 ⑥

< 信頼区間の下限値に標本平均
標本平均の分布によれば
 $50 - \mu$
Excelの関数を用いれば
-1.95996

⑥標本平均の分布と Excel の関数を用いた正規分布の値から、母平均の最大値を算出

母平均の最大値 51.95996

標本平均が95%区間上限値にある場合 ⑦

< 信頼区間の上限値に標本平均
標本平均の分布によれば
 $50 - \mu$
Excelの関数を用いれば
1.959964

⑦標本平均が信頼区間の上限値に位置している場合を作図する

最小値 48.04004

母平均の最小値の推定 ⑧

⑧標本平均と Excel の関数を用いた正規分布の値から、母平均の最小値を算出

シート名 : 母分散が未知

◆ 課題

このツールを用いて例題の解を求め、区間推定のプロセスを確認する。

標本平均に基づいて母平均の区間推定を行う・・・母分散が未知の場合



シート名 : 母比率の区間推定

得られた標本データに基づいて、母比率の信頼区間の推定を行う。

このツールを用いて例題の解を求め、区間推定のプロセスを確認する。

標本比率に基づいて母比率の区間推定を行う



(4) 1 標本の平均値に対する検定 (母分散未知)

統計検定の準備 (8)

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : 1 標本の平均値の検定.XLS

シート名 : データの分布を調べる → 問題の設定と検定

◆目的

得られた 1 標本データに基づいて、母平均の仮説検定を行う (母分散は未知である)。

◆課題

このツールを用いて例題の解を求め、仮説検定の考え方、計算手順に対する理解を確実なものとする。なお、検定を行う前に、「データの分布を調べる」シートでデータの概要について確認しておくこと。

◆手順 1 データの設定と分布の参照

A	B	C	D	E	F	G	H	I	
データNo.	データ	帰無仮説		標本No.		差	データの入力		
1	11	10		要約統計量	データ数		①		
2	7				平均値				
3	10				不偏分散				③
② 4	10				標準偏差				
5	9			観測値の分布(平均値と標準偏差)				作図	
6	14								
7	9			観測値の分布					
8	8								
9	8								
10	9								
11	11								
12	14								

①入力欄のクリア

「データの入力」ボタンをクリックすると、現在入力されている全てのデータが削除され、新たなデータの設定が可能となる。

②データの設定

②の欄に以下のデータを設定する

*データ No. (1 からの連番)

*データ

*帰無仮説

③要約統計量の算出

入力したデータに関する以下の要約統計量を算出し、③の欄に設定する。

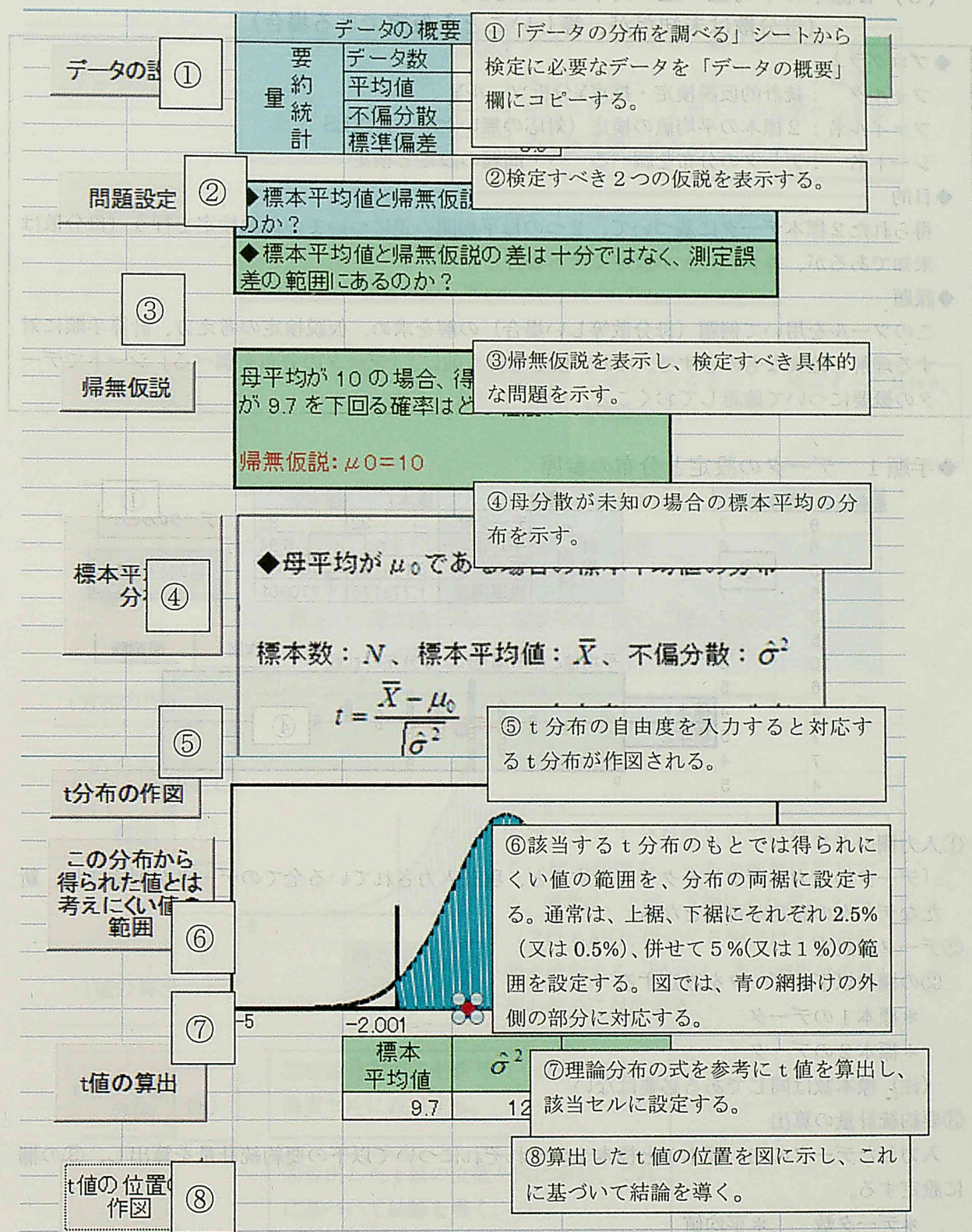
*データ数 *平均値

*不偏分散 *不偏標準偏差

④分布の作図

入力したデータの分布及び帰無仮説の位置を④の欄に作図する。分布の位置 (平均値) と帰無仮説の位置とそのずれの大きさを確認しておく。

◆手順2 検定の実行



(5) 2標本の平均値の差に対する仮説検定

(母分散は未知だが、等しいことを仮定できる場合)

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : 2標本の平均値の検定 (対応の無いケース) .XLS

シート名 : データの分布を調べる → 問題の設定と検定

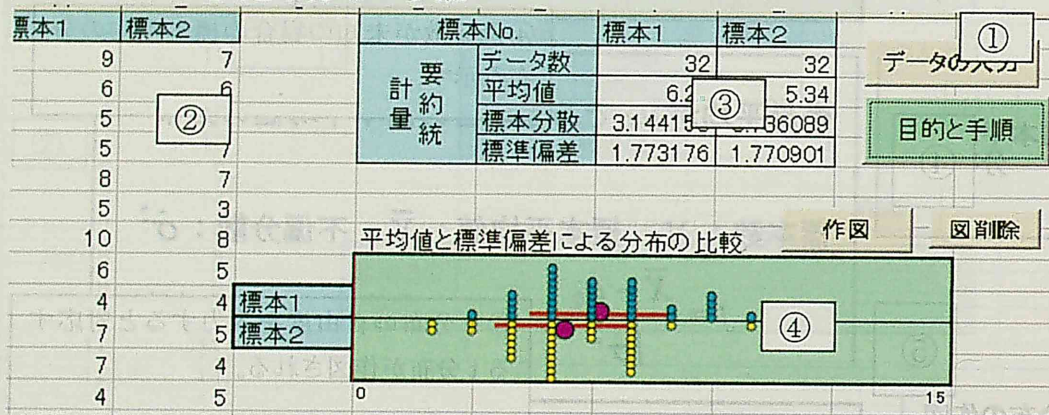
◆目的

得られた2標本データに基づいて、2つの母平均間の差についての仮説検定を行う (母分散は未知であるが、等しいことが仮定できる場合)。

◆課題

このツールを用いて例題 (母分散等しい場合) の解を求め、仮説検定の考え方、計算手順に対する理解を確実なものとする。なお、検定を行う前に、「データの分布を調べる」シートでデータの概要について確認しておくこと。

◆手順1 データの設定と分布の参照



①入力欄のクリア

「データの入力」ボタンをクリックすると、現在入力されている全てのデータが削除され、新たなデータの設定が可能となる。

②データの設定

②の欄に以下のデータを設定する

* 標本1のデータ

* 標本2のデータ

(注) 標本数は同じである必要はない

③要約統計量の算出

入力したデータの内、標本1と標本2のそれぞれについて以下の要約統計量を算出し、③の欄に設定する。

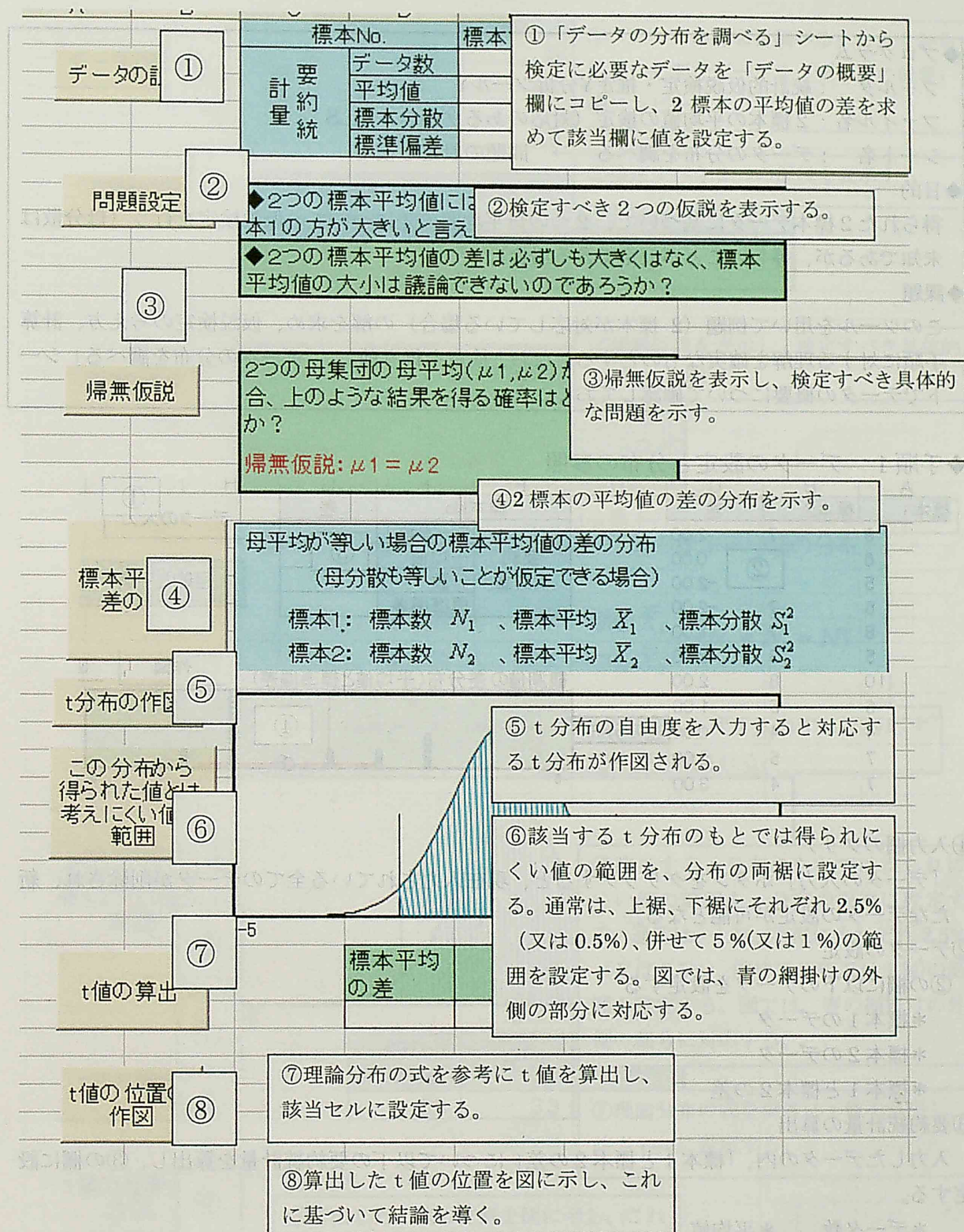
* データ数 * 平均値

* 標本分散 (Excel の関数は Varp を用いる) * 標本標準偏差

④分布の作図

標本1と標本2の分布を作図する。2つの分布がどの程度離れているかを確認しておく。

◆手順2 検定の実行



(6) 2標本の平均値の差に対する仮説検定

(2標本が対応している場合)

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : 2標本の平均値の検定 (対応のあるケース) .XLS

シート名 : データの分布を調べる → 問題の設定と検定

◆目的

得られた2標本データに基づいて、2つの母平均間の差についての仮説検定を行う (母分散は未知であるが、等しいことが仮定できる場合)。

◆課題

このツールを用いて例題 (2標本が対応している場合) の解を求め、仮説検定の考え方、計算手順に対する理解を確実なものとする。なお、検定を行う前に、「データの分布を調べる」シートでデータの概要について確認しておくこと。

◆手順1 データの設定と分布の参照

A	B	C	D	E	F	G	H	I
標本1	標本2	差		標本No.	差		データの①	
9	7	2.00		要約統計量	データ数	③	データの①	
6	②	0.00			平均値		目的と手順	
5		-2.00			分散			
5	7	-2.00			標準偏差			
8	7	1.00						
5	3	2.00						
10	8	2.00						
6	5	1.00						
4	4	0.00	差の分布	観測値の差分布 (平均値と標準偏差)	④	作図		
7	5	2.00						
7	4	3.00						

①入力欄のクリア

「データの入力」ボタンをクリックすると、現在入力されている全てのデータが削除され、新たなデータの設定が可能となる。

②データの設定

②の欄に以下のデータを設定する

*標本1のデータ

*標本2のデータ

*標本1と標本2の差

③要約統計量の算出

入力したデータの内、「標本1と標本2の差」について以下の要約統計量を算出し、③の欄に設定する。

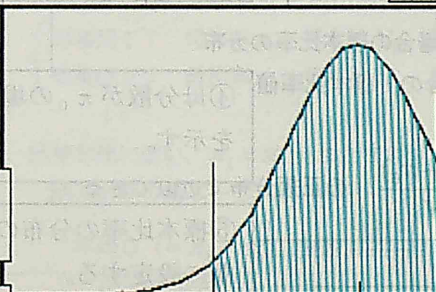
*データ数 *平均値

*不偏分散 *不偏標準偏差

④分布の作図

標本1と標本2の差の分布、及び「0」の位置を④の欄に作図する。分布の位置 (平均値) が「0」からどの程度離れているかを確認しておく。

◆手順2 検定の実行

		標本No.	差					
データの設 計 ①	要約統計	データ数	32	①「データの分布を調べる」シートから検定に必要なデータを「データの概要」欄にコピーする。				
		平均値	0.9					
		不偏分散	3.3					
		不偏標準偏差	1.8					
問題設定 ②	<p>2標本の差の平均値は0よりも一方が観測値は大きいと言えるのであろうか？</p> <p>◆2標本の差の平均値は殆ど0と見なすことができ、標本間の大小は議論できないのであろうか？</p>			②検定すべき2つの仮説を表示する。				
帰無仮説 ③	<p>2つの母集団の差の母平均が0のような結果を得る確率はどの程度か？</p> <p>帰無仮説: $\mu_0 = \mu_1 - \mu_2$</p>			③帰無仮説を表示し、検定すべき具体的な問題を示す。				
2標本の平均値 ④	<p>④2標本の差の平均値の分布を示す。</p> <p>差の分布の母平均が0である場合の、差の標本平均値の分布</p> <p> $\left\{ \begin{array}{l} \text{標本1: 標本数 } N_1, \text{ 標本平均値 } \bar{X}_1 \\ \text{標本2: 標本数 } N_2, \text{ 標本平均値 } \bar{X}_2 \end{array} \right. \quad (N_1 = N_2 = N)$ </p>							
t分布の作図 ⑤				⑤ t分布の自由度を入力すると対応するt分布が作図される。				
この分布から得られた値とは考えにくい値の範囲 ⑥	<p>⑥該当するt分布のもとでは得られにくい値の範囲を、分布の両裾に設定する。通常は、上裾、下裾にそれぞれ2.5% (又は0.5%)、併せて5% (又は1%)の範囲を設定する。図では、青の網掛けの外側の部分に対応する。</p>							
t値の算出 ⑦	<table border="1"> <thead> <tr> <th>差の平均値</th> <th>$\hat{\sigma}^2$</th> </tr> </thead> <tbody> <tr> <td>0.9</td> <td>3.3</td> </tr> </tbody> </table>			差の平均値	$\hat{\sigma}^2$	0.9	3.3	⑦理論分布の式を参考にt値を算出し、該当セルに設定する。
差の平均値	$\hat{\sigma}^2$							
0.9	3.3							
t値の位置の作図 ⑧	<p>⑧算出したt値の位置を図に示し、これに基づいて結論を導く。</p>							

(7) 1 標本の母比率に対する仮説検定

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : 標本比率の検定.XLS

シート名 : 1 標本の比率の検定

◆目的 得られた標本データに基づいて、母比率についての仮説検定を行う。

◆課題

このツールを用いて例題の解を求め、仮説検定の考え方、計算手順に対する理解を確実なものとする。

◆手順 検定の実行

A	B	C	D	E	F
		標本データ			帰無仮説
データの	①	要約	データ数	100	母比率
		統計量	標本比率	0.45	0.50

①データ数、標本比率、及び母比率に関する帰無仮説の値を設定する。

問題設定

②

◆標本比率と帰無仮説(母比率)の差は十分に大きいと言えるのか?
◆標本比率と帰無仮説(母比率)の差は十分に測定誤差の範囲にあるのか?

②検定すべき2つの仮説を表示する。

帰無仮説

③

母比率が0.6の場合、得られる標本比率が0を下回る確率ほどの程度か?
帰無仮説: $\pi_0 = 0.6$

③帰無仮説を表示し、検定すべき具体的な問題を示す。

標本比率の分布

④

◆母比率が π_0 である場合の標本比率の分布

標本数: N の場合の、標本比率値

平均 π_0 、分散 $\frac{\pi_0(1-\pi_0)}{N}$ の正規分布で近似できる

④母分散が π_0 の場合の標本比率の分布を示す。

平均	分散
0.50	0.0025

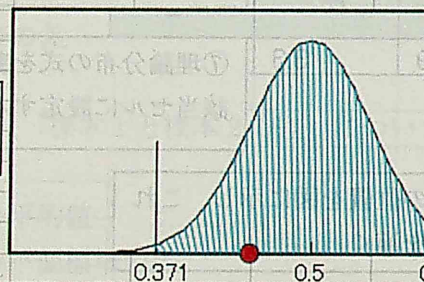
⑤標本比率の分布の平均値、分散を該当欄に設定する。

正規分布図

⑤

この分布
得られた値では
考えにくい値の
範囲

⑥



⑥該当する正規分布のもとでは得られにくい値の範囲を、分布の両裾に設定する。通常は、上裾、下裾にそれぞれ2.5%(又は0.5%)、併せて5%(又は1%)の範囲を設定する。図では、青の網掛けの外側の部分に対応する。

標本比率の位置

⑦

⑦得られている標本比率の位置を図に示し、これに基づいて結論を導く。

(8) 2標本の母比率の差に対する仮説検定

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : 標本比率の検定.XLS

シート名 : 2標本の比率の差の検定

◆目的 : 得られた2標本データに基づいて、母比率の差についての仮説検定を行う。

◆課題

このツールを用いて例題の解を求め、仮説検定の考え方、計算手順に対する理解を確実なものとする。

◆手順 検定の実行

A	B	C	D	E	F
		要約統計量	標本	第1標本	第2標本
データの	①		データ数	100	50
			標本比率	0.5	0.4

① 2つの標本について、データ数と標本比率を設定し、2つの母比率の差に関する仮説値を設定する。仮説値は通常0とする。

問題設定	②
	2つの標本比率の差は十分に大きいと言
	2つの標本比率の差は十分ではなく、測定誤差の範囲にあるのか？

② 検定すべき2つの仮説を表示する。

帰無	③
	母比率の差が0の場合、得られる標本比率の差が0.05を上回る確率はどの程度か？
	帰無仮説: $\pi_1 - \pi_2 = 0$

③ 帰無仮説を表示し、検定すべき具体的な問題を示す。

標本比率の差の分布	④				
	◆標本比率の差の分布 ($\pi_1 = \pi_2$ の場合) <ul style="list-style-type: none"> 母集団1 : 母比率 π_1、標本数 N_1、標本比率 P_1 母集団2 : 母比率 π_2、標本数 N_2、標本比率 P_2 ↓ <p>帰無仮説に従い、$\pi_1 = \pi_2 = \pi$ とすれば、標本比率の差 $P_1 - P_2$ の分布は、</p> <p>平均0 ($= \pi_1 - \pi_2$)、分散 $\frac{\pi(1-\pi)}{N}$</p>				
	<table border="1"> <tr> <td>π</td> </tr> <tr> <td></td> </tr> <tr> <td>平均</td> <td>分散</td> </tr> </table>	π		平均	分散
π					
平均	分散				

④ 2つの標本比率の差の分布を示す。

正規分布	⑤
	この分布から得られた値と、えにくい値の範囲

⑤ 2つの標本を込みした標本比率、帰無仮説が正しい場合の2標本の比率の差の分布について、その平均と分散の値を設定する。

標本比率の差の位置	⑥

⑥ 該当する正規分布のもとでは得られにくい値の範囲を、分布の両裾に設定する。通常は、上裾、下裾にそれぞれ2.5% (又は0.5%)、併せて5% (又は1%) の範囲を設定する。図では、青の網掛けの外側の部分に対応する。

標本比率の差の位置	⑦
	⑦ 標本比率の差の位置を図に示し、これに基づいて結論を導く。

(9) χ^2 適合度検定

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : カイ 2 乗検定.XLS

シート名 : 適合度検定

◆目的

1つの質的変数の単純集計データに基づいて、その構成比率に関するモデルへの適合度を χ^2 適合度検定によって調べる。

◆課題

このツールを用いて上記の例題の解を求め、適合度検定の基本的な考え方とその計算手順に対する理解を確実なものとする。

◆手順 検定の実行

カイ自乗の適合度検定

適合性の検定

準備

①

カテゴリ数

開始

①分析対象の質的変数のカテゴリ数を入力する。これにより、シート上に指定したカテゴリ数の単純集計表が用意されるので、観測度数と帰無仮説である各カテゴリの期待比率（モデル比率）を入力する。

②

計算開始

カテゴリ	カテゴリ-1	カテゴリ-2	カテ
観測度数	15	22	
モデル比率	0.2	0.2	
モデル度数(期待値)	16.8	16.8	
誤差=観測-期待値	-1.8	5.2	
誤差 ² ÷期待値	0.192857	1.609524	0.0

②モデルが正しい場合の期待観測度数、観測度数と期待度数との差、及び差の 2 乗を期待度数で割った値とその合計（カイ 2 乗値）が計算される。

③

モデルが正しい場合の観測度数と期待度数の和の分布

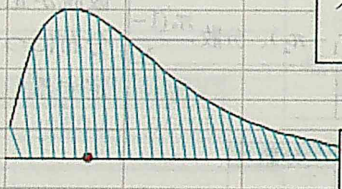
カイ自乗分布
自由度

4

③帰無仮説（モデル）が正しい場合の、カイ 2 乗値の分布が作図される。

④

標本カイ自乗値



棄却域と採択

⑤

観測値 2.547619 < 上限値 13.276

④実際に観測されたカイ 2 乗値の位置が作図される。

⑤帰無仮説が正しい場合には観測されにくいカイ 2 乗の範囲が、設定した有意水準に対応して与えられる。この範囲と観測カイ 2 乗値との関係から、最終的な結論を導く。

(10) 相関係数の検定

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : 相関係数の検定.XLS

シート名 : データの分布を調べる → 問題の設定と検定

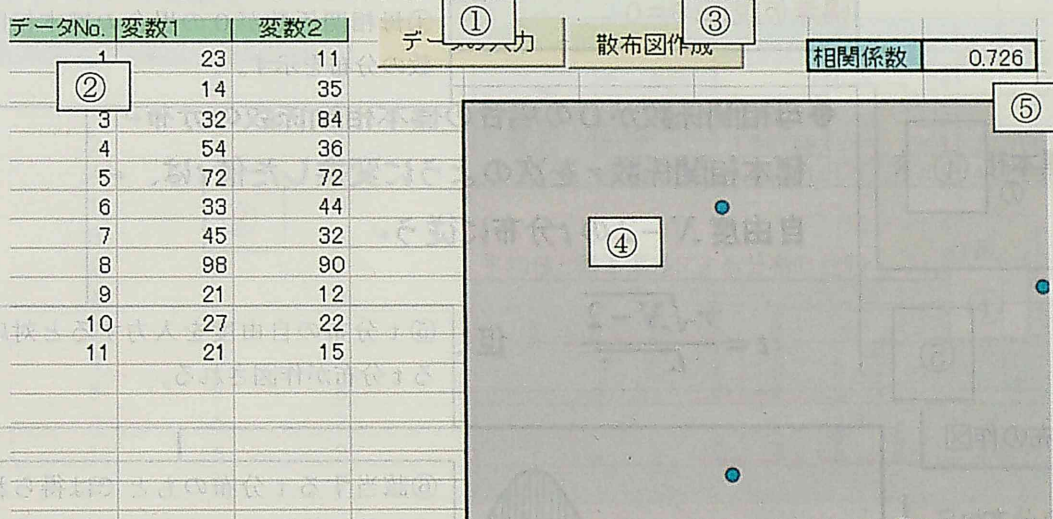
◆目的

2変数間に有意な関連があるか否かを、標本相関係数値に基づいて検定する。

◆課題

このツールを用いて前記の例題の解を求め、相関係数検定の基本的な考え方とその計算手順に対する理解を確実なものとする。

◆手順1 データの設定と散布図の参照



①入力欄のクリア

「データの入力」ボタンをクリックすると、現在入力されている全てのデータが削除され、新たなデータの設定が可能となる。

②データの設定

②の欄に以下のデータを設定する

*データ No. (1からの連番)

*変数1のデータ

*変数2のデータ

③散布図の作成と相関係数の算出

入力したデータに関する散布図を欄④に作図し、相関係数を欄⑤に示す。

◆手順2 検定の実行

データの①

標本相関係数	
データ数	11
相関係数	0.726

問題設定②

③
帰無仮説

④
標本相関係数の分布

⑤
t分布の作図

⑥
この分布から得られた値とは考えにくい値範囲

⑦
t値の算出

⑧
t値の位置作図

①「データの分布を調べる」シートから検定に必要なデータを「標本相関係数」欄にコピーする。また、帰無仮説として、母相関係数数値に0を設定する。

②検定すべき2つの仮説を表示する。

③帰無仮説を表示し、検定すべき具体的な問題を示す。

④母相関係数が0の場合の標本相関係数の分布を示す。

◆母相関係数が0の場合の標本相関係数の分布

標本相関係数 r を次のように変換した値 t は、自由度 $N-2$ の t 分布に従う

$$t = \frac{r\sqrt{N-2}}{L}$$

但し

⑤ t 分布の自由度を入力すると対応する t 分布が作図される。

⑥該当する t 分布のもとでは得られにくい値の範囲を、分布の両裾に設定する。通常は、上裾、下裾にそれぞれ2.5% (又は0.5%)、併せて5%(又は1%)の範囲を設定する。図では、青の網掛けの外側の部分に対応する。

標本相関係数	t値
0.7	3.16667

⑦理論分布の式を参考に t 値を算出し、該当セルに設定する。

⑧算出した t 値の位置を図に示し、これに基づいて結論を導く。

(11) 一要因の分散分析

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : 一要因分散分析 (被験者間) .XLS

シート名 : データの分布を調べる → 多重比較

◆目的

異なった条件下で得られた複数個の標本データの分布に有意な違いがあるか否かを分散分析 (F 検定) によって調べる。

◆課題

例題について、条件間の変動 (級間平方和)、条件内での変動 (級内平方和) の算出から分散分析表の作成までの処理を行い、最終的に F 検定によって結果を評価する。このツールを用いて、F 検定 (一要因分散分析) の基本的な考え方とその計算手順に対する理解を確実なものとする。

◆手順1 データの設定

①

水準-1	水準-2	水準-3
5	2	3
1	5	7
2	5	4
4		4
1		6
3		4
2	5	3
4	4	6
2	5	6
4	6	7
5	5	6
4	5	7
1	5	3
3	6	7

②

③

要約統計量	データ数	水準-1	水準-2	水準-3
データ数	24	24	24	24
平均値	3.208333	3.333	5.25	5.25
分散	1.581597	1.931	2.270833	2.270833
標準偏差	1.257616	1.135751	1.506928	1.506928

④

平均値と標準偏差による分布の比較

①データ入力欄をクリアする

②解析データを②の欄に設定する

③統計量を全て算出する

④水準ごとのデータ分布を作成し、比較しておく

◆手順2

①		処理	水準-1	水準-2	水準-3	平均
データの設定	各水準、全体の平均値を求める		3.21	4.71	5.25	4.39
	(水準平均 - 全平均)		-1.18		0.86	
	(水準平均 - 全平均) ²		1.39		0.74	
	(水準平均 - 全平均) ² × データ数		33.45	2.45	17.80	
処理内容						合計
						53.69

①データの設定シートで入力したデータを設定する

②要因の効果を「処理内容ボタン」を参考に計算する

◆手順3

①

データの設定

処理内容

種別	水準-1	水準-2	水準-3
	5	2	3
	1	5	7
	2	5	4
	4	3	4
	1	4	6
	3	4	4
	2	5	3
	4	4	6

②

①データの設定シートで入力したデータを設定する

②誤差の大きさを「処理内容ボタン」を参考に計算する

◆手順4

分散分析表

①

要因	平方和	自由度	平均平方和	F値	検定結果
因子	53.69444	2	26.84722	15.00979	1%有意
誤差	123.4167	69	1.788647		
合計	177.1111	71			

②

F分布作図

境界値設定

③

図削除

上限値 4.926671

観測F値の位置

④

- ①平方和、自由度、平均平方和、F値を算出し、分散分析表を完成させる
- ②要因効果が全く無い場合のF値の分布（F分布）を描く
- ③有意水準に対応した境界値を設定する
- ④観測されたF値を図の中に位置づけ、検定結果を得る

◆手順5 多重比較

①

比較対象	比較結果	検定結果
因子全 - 誤差	有意	1%有意
因子全 - 因子全	有意	1%有意
因子全 - 誤差	有意	1%有意
因子全 - 因子全	有意	1%有意

(12) 二要因の分散分析

◆プログラム

フォルダ : 統計的仮説検定・推定・分析ツール

ファイル名 : 二要因分散分析.XLS

シート名 : データの分布を調べる → 多重比較

◆目的

2つの要因のもとで観測された複数個の標本データの分布に有意な違いがあるか否か、つまり要因効果、及び交互作用は有意性を分散分析（F検定）によって調べる。

◆課題

要因の主効果、及び交互作用項の効果、さらに誤差の変動分を算出した後、それらのデータを元に分散分析表を作成し、最終的にF検定によって結果を評価する。このツールを用いて、(二) 二要因分散分析の基本的な考え方とその計算手順に対する理解を確かなものとする。

◆手順1 データの設定

データの入力	第1要因	第2要因			第1要因	第2要因
		第1水準	第2水準	第3水準		
実験計画 要因 水準数 第1要因 3 第2要因 3 繰り返し数 5 作図 図削除	第1水準	1	2	3	水準1	③
		2	1	4	水準2	
		3	2	5	水準3	
	第2水準	4	3	6	水準1	
		5	4	7	水準2	
		6	5	8	水準3	
	第3水準	7	6	9	水準1	
		8	7	10	水準2	
		9	8	11	水準3	
		10	9	12	水準1	
		11	10	13	水準2	
		12	11	14	水準3	

①データ入力欄をクリアする

②解析データを②の欄に設定する

④水準ごとのデータ分布を作成し、比較しておく

◆手順2

データの設定	処理	水準-1	水準-2	水準-3	平均	合計
処理の内容	水準別平均、全平均を求める	2.533	4.733	8.067	5.111	
	(水準平均-全平均)	-2.578	-0.378	2.956		
	(水準平均-全平均) ²	6.64	0.14	8.74		
	(水準平均-全平均) ² ×データ数	99.674	2.14074	131.0296		232.8444
作図	第1要因	第1水準	第2水準	第3水準		
	第1	1	2	3		
	第2	2	1	4		
	第3	3	2	5		

①データの設定シートで入力したデータを設定する

②第1要因の効果を「処理内容」ボタンを参考に計算する

◆手順3

①	データの設定	水準別平均、全平均を求める	水準-1	水準-2	水準-3	平均	
	処理の内容	(水準平均-全平均)	-1.644	0.08	556		
	作図	(水準平均-全平均) ²	2.704	0.005	2.420		
		(水準平均-全平均) ² ×データ数	40.56	0.12	36.30		合計 76.98
			第1要因		第2要因		
			第1水準		第2水準	第3水準	
			1		2	3	

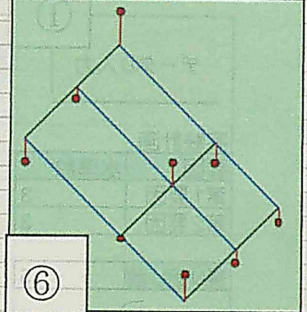
①データの設定シートで入力したデータを設定する

②第2要因の効果を「処理内容」ボタンを参考に計算する

◆手順4

①	データの設定	主効果による予測値を求める	第1要因	第2要因			主効果	
	処理の内容		第1水準	第2水準	第3水準			
			0.889	4.089	-2.577778			
			3.089	6.289	-0.377778			
Tukey Graph	各要因-水準ごとの平均値を求める		第1水準	第2水準	第3水準			
			1.800	3.600				
			3.000	5.600				
			5.600	7.800	10.800			
図削除	主効果による予測値と平均値との残差を求める		第1水準	第2水準	第3水準			
			0.911	-0.422				
			-0.089	0.778				
			-0.822	-0.356				
	交互作用の大きさを見積もる		第1水準	第2水準	第3水準			
			4.150617	0.89135802	1.19506			
			0.039506	3.02469136	2.37283			
			3.380247	0.63209877	6.93580			
			合計					22.62222

< Tukey Graph >



①データを設定する

②主効果による予測値を求める

③要因の組み合わせごとに観測値の平均を求める

④平均値と主効果による予測値との差を求める

⑤差の2乗に観測数をかけ、全て加えて交互作用項の変動分を求める

⑥交互作用項の大きさを図で確認する

◆手順5

データの設定		<観測値>				<水準の組み合わせにおける平均値>			
処理内容	① 平均からの偏差を求める	第1要因		第2要因		第1要因		第2要因	
		第1水準	第2水準	第3水準	第3水準	第1水準	第2水準	第3水準	第3水準
		第1水準	1	2	3	第1水準	1.800	2.200	3.600
		第1水準	2	1	4	第2水準	3.000	5.600	5.600
		第1水準	1	3	2	第3水準	5.600	7.800	10.800
		第1水準	3	2	5				
		第1水準	2	3	4				
		第2水準	3	5	7				
		第2水準	2	6	5				
		第2水準	3	4	4				
		第2水準	4	7	5				
		第2水準	3	6	7				
		第3水準	5	7	9				
		第3水準	6	8	10				
		第3水準	7	9	11				

- ①データの設定シートで入力したデータを設定する
- ②誤差の大きさを「処理内容」ボタンを参考に計算する

◆手順6

分散分析表		①				①
処理内容	要因	平方和	自由度	平方和	F値	検定結果
	第1要因					
	第2要因					
	交互作用					
	誤差					
	全体					
F分布作図		③				③
境界値の設定		④				④
観測F値の位置		⑤				⑤
②	第1要因					
	第2要因					
	交互作用					

- ①平方和、自由度、平均平方和、F値を算出し、分散分析表を完成させる
- ②検定を行う要因を指定する
- ③要因効果が全く無い場合のF値の分布（F分布）を描く
- ④有意水準に対応した境界値を設定する
- ⑤観測されたF値を図の中に位置づけ、検定結果を得る

(13) クロス表の関連度についての χ^2 検定

◆プログラム

フォルダ : 統計的仮説検定・推定分析ツール

ファイル名 : カイ2乗検定.XLS

シート名 : 独立性の検定

◆目的

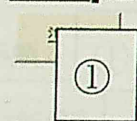
クロス表の関連度を χ^2 検定によって評価する。

◆課題

表 7-6 のクロス表について検定を行い、 χ^2 検定の基本的な考え方とその計算手順に対する理解を確実なものとする。

◆検定の実行

2変数が独立の場合



2変数が独立の場合

誤差 = 観測値 - 期待値

準備

手順

③

2変数が独立の場合

④

誤差 = 観測値 - 期待値

①準備ボタンをクリックするとクロス表の定義フォームが表示されるので、変数名とカテゴリ数を入力する

→ 以下のクロス表の水色欄を数値で埋め
< 観測されたクロス表 >

変数A	カテゴリ	変数B				周辺度数
	カテゴリ	カテゴリ-1	カテゴリ-2	カテゴリ-3	カテゴリ-4	
変数A	カテゴリ-1	30	20	10	5	65
	カテゴリ-2	20	20	20	20	80
	カテゴリ-3	10	20	30	40	100
	周辺度数	60	60	60	65	245

< 2変数が独立の場合の期待値 >

変数A	カテゴリ	変数B				周辺度数
	カテゴリ	カテゴリ-1	カテゴリ-2	カテゴリ-3	カテゴリ-4	
変数A	カテゴリ-1	15.9	15.9	15.9	17.2	65
	カテゴリ-2	19.6	19.6	19.6	21.2	80
	カテゴリ-3	24.5	24.5	24.5	26.5	100
	周辺度数	60	60	60	65	245

③2変数が独立な場合の期待度数が表示される。

< 観測度数と期待度数との差(誤差) >

変数A	カテゴリ	変数B				周辺度数
	カテゴリ	カテゴリ-1	カテゴリ-2	カテゴリ-3	カテゴリ-4	
変数A	カテゴリ-1	14.1	4.1	-5.9	-10.2	65
	カテゴリ-2	0.4	0.4	0.4	-1.2	80
	カテゴリ-3	-14.5	-4.5	5.5	13.5	100
	周辺度数	60	60	60	65	245

④観測度数と独立な場合の期待度数との差が計算される。

⑤
誤差2乗

< 誤差の2乗 ÷ 期待値 >

変数 A	カテゴリ	カテゴリ-1	カテゴリ-2	カテゴリ-3	合計
	カテゴリ-1	12.5	1.0	0.0	13.5
	カテゴリ-2	0.0	0.0	0.1	0.1
	カテゴリ-3	8.6	0.8	1.2	10.6
	合計	21.0	1.9	3.4	26.3

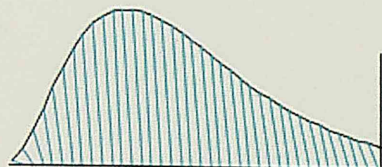
⑤ 観測度数と期待度数との差の2乗を期待度数で割った値が計算され、その全セル合計（カイ2乗値）が右下隅に表示される。

⑥
モデルが適合している場合の誤差2乗 ÷ 期待値の和の半分

自由度 6 のカイ自乗分布

⑥ 対応する自由度を持つカイ2乗分布を作図する。

⑦
標本カイ



⑦ クロス表から計算されたカイ2乗値の位置が作図される。

⑧
棄却域と採択

上限値 12.59159 < 観測値 41.96893

⑧ 変数が独立である場合には、観測されにくいカイ2乗の範囲が、設定した有意水準に対応して与えられる。この範囲と観測カイ2乗値との関係から、最終的な結論を導く。

**CD-ROMは カウンターに
お申し込みください。**